DOI:
10.1080/13506285.2017.1336141

Document Version
Peer reviewed version

Visual Cognition, doi:10.1080/13506285.2017.1336141

A predictive coding model of gaze shifts and the underlying neurophysiology

M. W. Spratling
King's College London, Department of Informatics, London. UK. michael.spratling@kcl.ac.uk

Abstract

A comprehensive model of gaze control must account for a number of empirical observations at both the behavioural and neurophysiological levels. The computational model presented in this article can simulate the coordinated movements of the eye, head, and body required to perform horizontal gaze shifts. In doing so it reproduces the predictable relationships between the movements performed by these different degrees of freedom (DOFs) in the primate. The model also accounts for the saccadic undershoot that accompanies large gaze shifts in the biological visual system. It can also account for our perception of a stable external world despite frequent gaze shifts and the ability to perform accurate memory-guided and double-step saccades. The proposed model also simulates peri-saccadic compression: the mis-localisation of a briefly presented visual stimulus towards the location that is the target for a saccade. At the neurophysiological level, the proposed model is consistent with the existence of cortical neurons tuned to the retinal, head-centred, body-centred, and world-centred locations of visual stimuli and cortical neurons that have gain-modulated responses to visual stimuli. Finally, the model also successfully accounts for peri-saccadic receptive field (RF) remapping which results in reduced responses to stimuli in the current RF location and an increased sensitivity to stimuli appearing at the location that will be occupied by the RF after the saccade. The proposed model thus offers a unified explanation for this seemingly diverse range of phenomena. Furthermore, as the proposed model is an implementation of the predictive coding theory, it offers a single computational explanation for these phenomena and relates gaze shifts to a wider framework for understanding cortical function.

Keywords: eye movements; saccades; peri-saccadic compression; gain-modulation; receptive field remapping; sensory-motor coordination; neural networks; basis functions; action planning

1 Introduction

Movements that change the direction of gaze (eye movements or coordinated movements of the eye, head and body) are one of the most frequent actions performed by humans and many other species. As a result, such movements have been extensively studied, using both psychophysics and electrophysiology. These experiments have revealed a wide range of different phenomena associated with gaze shifts (Hamker et al., 2011; McCluskey and Cullen, 2007; Melcher and Colby, 2008; Wurtz, 2008), including: predictable relationships between eye, head and body movements; typical patterns of saccade accuracy; the ability to accurately perform memory-guided and double-step saccades; inaccuracies in perceiving the location of a visual stimulus that appears briefly around the time of a saccade (peri-saccadic shifts and compression); suppression of awareness of visual blur during saccades (saccadic suppression); the perception of a stable external world despite frequent gaze shifts; a failure to detect visual target location changes made during a saccade (suppression of displacement); the existence of cortical neurons tuned to the retinal, head-centred, body-centred, and world-centred locations of visual stimuli; neurons with visual responses that are gain-modulated by postural information; and, neurons that show predictive responses to stimuli that will appear in their receptive field (RF) following a saccade (RF remapping).

While there are many existing computational models that offer explanations of these phenomena individually (*e.g.*, Dominey and Arbib, 1992; Droulez and Berthoz, 1991; Krommenhoek et al., 1993; Law et al., 2013; Mender, 2014; Mender and Stringer, 2015; Niemeier et al., 2003; Pola, 2011; Pouget et al., 2002; Pouget and Sejnowski, 1997; Pouget and Snyder, 2000; Quaia et al., 1998; Saeb et al., 2011; Salinas and Abbott, 1995, 1996; Salinas and Sejnowski, 2001; Schneegans, S.and Schöner, 2012; Smith and Crawford, 2001, 2005; Tweed, 1997; Weber et al., 2007; Xing and Andersen, 2000a,b; Zipser and Andersen, 1988), or in twos or threes (Hamker et al., 2008; Mohsenzadeh et al., 2016; Ziesche and Hamker, 2011, 2014), there is currently a lack of a more comprehensive model that can account for many gaze-related phenomena. This article proposes such a comprehensive model. The proposed computational model can account for all the empirical data listed above, except for peri-saccadic shifts, saccadic suppression, and suppression of displacement. The proposed model is based on a particular implementation of the predictive coding theory of cortical function (Bubic et al., 2010; Clark, 2013; Friston and Kiebel, 2009; Huang and Rao, 2011; Rao and Ballard, 1999). Previous work has already applied predictive coding to modelling action (Adams et al., 2013; Friston et al., 2012, 2011, 2010; Perrinet et al., 2014). This prior work emphasises

that actions can be viewed as a means of reducing prediction error by changing sensory data to make it consistent with the expected or desired sensory input. In these previous predictive coding models of action, fulfilling expectations about predicted sensory inputs is achieved by minimising the error between those expectations and the actual sensory input. However, this previous work does not simulate psychophysical or neurophysiological data. The emphasis here is on explaining such data, and on exploring how predictive coding mechanisms can be used to plan actions, rather than being concerned with modelling how actions are executed once planned.[a] The actions planned by the proposed model could well be executed through error minimisation, using predictive coding, as advocated in the previous work (Adams et al., 2013; Friston et al., 2012, 2011, 2010; Perrinet et al., 2014), but this is not considered here.

The proposed model uses population codes to represent different sensory and motor variables, consistent with the principle set out by van Hemmen and Schwartz (2008). The relationships between these variables are encoded in the weights of a predictive coding neural network. This enables the network to perform mappings between the different variables, predicting any missing values. These predictions of missing values can be used to plan movements by predicting the position of the eye, neck, or torso required to bring a visual target to a particular location on the retina. The predictions generated by the network can also be used to estimate the visual input expected after a change in gaze direction. While it is possible to encode the relationships between all the sensory and motor variables using a single network, the number of neurons required increases exponentially with the number of variables, and is unlikely to be a tractable solution for the brain (Deneve and Pouget, 2003; Pouget and Sejnowski, 1997). To resolve this issue the overall problem is divided across multiple sub-networks. This subdivision requires internal variables (also population coded) that facilitate the coordination of the sub-networks. Natural candidates for these internal variables are representations of visual space that are increasingly invariant to movements: representations in head-centred, body-centred, and world-centred coordinates. These gaze-invariant representations allow accurate eye movements to be performed to previously seen locations even when intervening movements have been made. The use of population coding results in the model exhibiting patterns of eye, head and body movements that are consistent with those observed in humans, it also results in saccadic undershoot for large gaze shifts that is also observed in the biological visual system. The relationship between different variables is encoded by neurons that have gain modulated responses. Using the model to predict the visual input expected after a gaze shift produces neurons that show predictive RF remapping responses. Furthermore, using the same neural network to make multiple predictions about different variable values leads naturally to an explanation of peri-saccadic compression.

2 Methods

2.1 Basis Function Networks

Basis function networks are a popular neural network architecture for performing sensory-sensory and sensory-motor coordination in robots (Kim et al., 2005; Marjanović et al., 1996; Meng and Lee, 2007, 2008; Molina-Vilaplana et al., 2004; Sun and Scassellati, 2005; Zhang et al., 2005) and as models of brain function (Chinellato et al., 2011; Deneve et al., 2001; Deneve and Pouget, 2003; Latham et al., 2003; Pouget et al., 2003, 2002; Pouget and Sejnowski, 1994, 1997; Pouget and Snyder, 2000; Salinas and Abbott, 1995; Salinas and Sejnowski, 2001; van Rossum and Renart, 2004; Weber et al., 2007). Basis function networks can approximate any linear or nonlinear mapping (Broomhead and Lowe, 1988; Park and Sandberg, 1991; Schilling et al., 2001), but for simplicity, a very simple linear example is shown in fig. 1. The basis function approach splits the problem into two sub-problems: a layer of basis function nodes, with nonlinear activation functions, encode possible combinations of sensory input signals, and a linear readout of the responses of these basis functions is used to produce the output.

The basis function network shown in fig. 1 could be interpreted as performing a sensory-sensory transformation that maps retinocentric coordinates to head-centred coordinates for a simple system with a one-dimensional retinotopic input and with one degree-of-freedom (DOF) for motor action which causes a shift in the retinotopic input (Deneve et al., 2001; Pouget et al., 2002; Pouget and Sejnowski, 1997; Pouget and Snyder, 2000; Weber and Wermter, 2007). In this case, the current value of a represents the position of a visual target on the retina, the value of b represents position of the eye, and the calculated value of c represents the estimate of the head-centred position of the target. This article uses basis function networks that are implemented using the PC/BC-DIM algorithm.

[a] The proposed model is concerned purely with kinematics and does not consider dynamics. It is assumed that the movements planned by the proposed model are carried out by brain circuitry that has not been explicitly modelled here. For example, by the cerebellum (Houk et al., 1996; Kawato, 1995), which appears to implement a closed-loop motor control system containing both forward and inverse models (Wolpert and Kawato, 1998; Wolpert et al., 1998). This split between action planning and execution is consistent with previous work modelling the biological basis of motor control (Flash and Sejnowski, 2001).

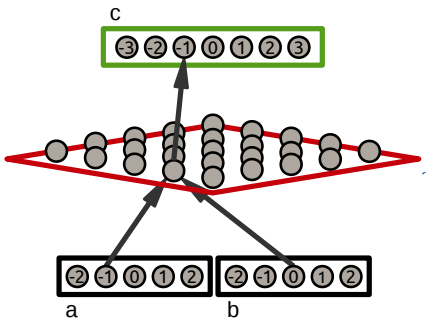


Figure 1: A simple basis function network for performing a mapping between two input variables (a and b) and an output variable (c). In this simple example, the mapping is linear, such that $c = a+b$. The values of the input variables are encoded by the activation patterns in two populations of neurons, and the value of the output is encoded by the firing of a third population of neurons. By using populating coding the inputs and outputs can represent any continuous value, rather than just discrete values as suggested by the figure. While representations of the current variable values are encoded by neural activity, the mapping between values is encoded in the connections (the "weights") between the neurons in the network. The mapping from the inputs to the output is mediated by a hidden layer of neurons, the basis function population. Each basis function node has weights that allow it to represent a possible combination of input values. Each neuron in the output layer has non-zero weights to those basis function neurons that represent the same output value (*e.g.*, an output neuron representing the value $c = -1$, would receive input from the basis function representing the combination of inputs $a = -1$ and $b = 0$ and would also receive input from the basis function representing the inputs $a = -2$ and $b = 1$, *etc.*).

2.2 The PC/BC-DIM Algorithm

PC/BC-DIM is a version of Predictive Coding (PC; Rao and Ballard, 1999) reformulated to make it compatible with Biased Competition (BC) theories of cortical function (Spratling, 2008a,b) and that is implemented using Divisive Input Modulation (DIM; Spratling et al., 2009) as the method for updating error and prediction neuron activations. DIM calculates reconstruction errors using division, which is in contrast to other implementations of PC that calculate reconstruction errors using subtraction (Huang and Rao, 2011; Spratling, 2017). PC/BC--DIM is a hierarchical neural network. Each level, or processing stage, in the hierarchy is implemented using the neural circuitry illustrated in fig. 2a. A single PC/BC-DIM processing stage thus consists of three separate neural populations. The behaviour of the neurons in these three populations is determined by the following equations:

$$\mathbf{r} = \mathbf{V}\mathbf{y} \tag{1}$$

$$\mathbf{e} = \mathbf{x} \oslash [\mathbf{r}]_{\epsilon_2} \tag{2}$$

$$\mathbf{y} \leftarrow [\mathbf{y}]_{\epsilon_1} \otimes \mathbf{W}\mathbf{e} \tag{3}$$

Where $\mathbf{x}$ is a (m by 1) vector of input activations, $\mathbf{e}$ is a (m by 1) vector of error neuron activations; $\mathbf{r}$ is a (m by 1) vector of reconstruction neuron activations; $\mathbf{y}$ is a (n by 1) vector of prediction neuron activations; $\mathbf{W}$ is a (n by m) matrix of feedforward synaptic weight values; $\mathbf{V}$ is a (m by n) matrix of feedback synaptic weight values; $[v]_\epsilon = \mathtt{max}(\epsilon, v)$; ϵ_1 and ϵ_2 are parameters; and $\oslash$ and $\otimes$ indicate element-wise division and multiplication respectively.

For all the experiments described in this paper ϵ_1 and ϵ_2 were given values of 1×10^{-6} and 1×10^{-4} respectively. Parameter ϵ_1 prevents prediction neurons becoming permanently non-responsive. It also sets each prediction neuron's baseline activity rate and controls the rate at which its activity increases when an input stimulus is presented within its RF. Parameter ϵ_2 prevents division-by-zero errors and determines the minimum strength that an input is required to have in order to effect prediction neuron response. As in all previous work with PC/BC-DIM, these parameters have been given small values compared to typical values of $\mathbf{y}$ and $\mathbf{x}$, and hence, have negligible effects on the steady-state activity of the network. The matrix $\mathbf{V}$ is equal to the transpose of the $\mathbf{W}$, but each column is normalised to have a maximum value of one. Hence, the feedforward and feedback weights are simply rescaled versions of each other. Given that the $\mathbf{V}$ weights are fixed to the $\mathbf{W}$ weights there is only one set of free parameters, $\mathbf{W}$, and references to the "synaptic weights" refer to the elements of $\mathbf{W}$. Here, as in previous work with PC/BC-DIM only non-negative weights, inputs, and activations are used. Initially the values of $\mathbf{y}$ are all set to zero, although random initialisation of the prediction node activations can also be used with little

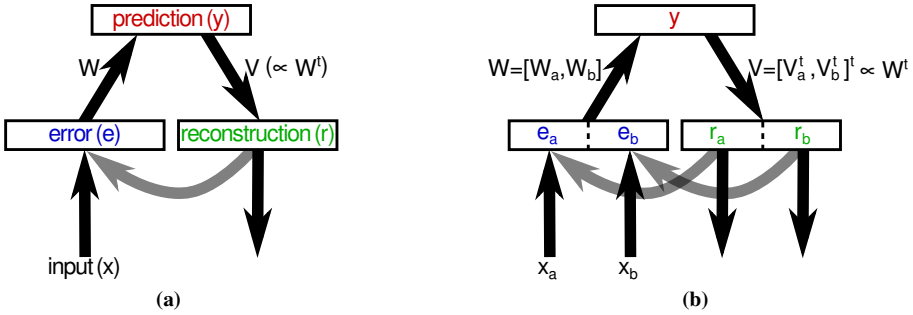


Figure 2: (a) A single processing stage in the PC/BC-DIM neural network architecture. Rectangles represent populations of neurons and arrows represent connections between those populations. The population of prediction neurons constitute a model of the input environment. Individual neurons represent distinct causes that can underlie the input (*i.e.*, latent variables). The belief that each cause explains the current input is encoded in the activation level, $\mathbf{y}$, and is used to reconstruct the expected input given the predicted causes. This reconstruction, $\mathbf{r}$, is calculated using a linear generative model (see eq. 1). Each column of the feedback weight matrix $\mathbf{V}$ represents a basis function, and the reconstruction is thus a linear combination of those basis vectors. Each element of the reconstruction is compared to the corresponding element of the actual input, $\mathbf{x}$, in order to calculate the residual error, $\mathbf{e}$, between the predicted input and the actual input (see eq. 2). The errors are subsequently used to update the predictions (via the feedforward weights $\mathbf{W}$, see eq. 3) in order to make them better able to account for the input, and hence, to reduce the error at subsequent iterations. The responses of the neurons in all three populations are updated iteratively to recursively calculate the values of $\mathbf{y}$, $\mathbf{r}$, and $\mathbf{e}$. The weights $\mathbf{V}$ are the transpose of the weights $\mathbf{W}$, but are normalised to so that the maximum value of each column is unity. The activations of the prediction neurons or the reconstruction neurons may be used as inputs to other PC/BC-DIM processing stages. The inputs to this processing stage may come from the prediction neurons of this or another processing stage, or the reconstruction neurons of another processing stage, or may be external, sensory-driven, signals. The inputs can also be a combination of any of the above. (b) When inputs come from multiple sources, it is convenient to consider the population of error neurons to be partitioned into sub-populations which receive these separate sources of input. As there is a one-to-one correspondence between error neurons and reconstruction neurons, this means that the reconstruction neuron population can be partitioned similarly.

influence on the results. Equations 1, 2 and 3 are then iteratively updated with the new values of $\mathbf{y}$ calculated by eq. 3 substituted into eq. 1 and eq. 3 to recursively calculate the neural activations.

There are multiple ways to create hierarchical PC/BC-DIM networks. One way is to use the prediction neurons in one processing stage, as (part of) the input to a subsequent stage (Spratling, 2008a, 2012c, 2016b). Another possibility is to use the the reconstruction neurons as inputs to another processing stage (Muhammad and Spratling, 2015; Spratling, 2016a,b). The latter approach will be used in this article to perform the simulations described in sect. 3.1. To perform simulations with a hierarchical PC/BC-DIM network, eqs. 1, 2 and 3 are evaluated for each processing stage in turn (starting from the lowest stage in the hierarchy), and this process is repeated to iteratively calculate the changing neural activations in each processing stage at each time-step.

When inputs come from multiple sources it is convenient to consider the vector of input signals, $\mathbf{x}$, the vector of error neuron activations, $\mathbf{e}$, and the vector of reconstruction neuron responses, $\mathbf{r}$, to be partitioned into multiple parts corresponding to these separate sources of input (see fig. 2b; Spratling, 2014b, 2016b). Each partition of the input will correspond to certain columns of $\mathbf{W}$ (and rows of $\mathbf{V}$). While it is conceptually convenient to think about separate partitions of the inputs, neural populations and synaptic weights, it does not in any way alter the mathematics of the model. In eqs. 1, 2 and 3, $\mathbf{x}$ is a concatenation of all partitions of the input, $\mathbf{e}$ and $\mathbf{r}$ represent the activations of all the error and reconstruction neurons; and $\mathbf{W}$ and $\mathbf{V}$ represent the synaptic weight values for all partitions.

In the PC/BC-DIM algorithm, the values of $\mathbf{y}$ represent predictions of the causes underlying the inputs to the network. The values of $\mathbf{r}$ represent the expected inputs given the predicted causes. The values of $\mathbf{e}$ represent the residual error between the reconstruction, $\mathbf{r}$, and the actual input, $\mathbf{x}$. The full range of possible causes that the network can represent are defined by the weights, $\mathbf{W}$ (and $\mathbf{V}$). Each row of $\mathbf{W}$ (which correspond to the weights targeting an individual prediction neuron) can be thought of as a "basis function" or "elementary component" or "preferred stimulus" or "dictionary element", and $\mathbf{W}$ as a whole can be thought of as a "basis", or "codebook",

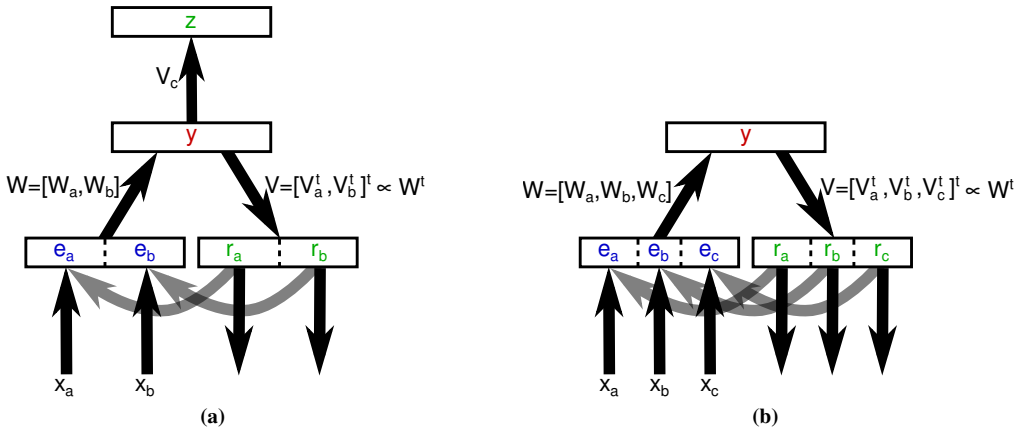


Figure 3: Methods of using PC/BC-DIM as a basis function network. The architectures shown here are analogous to that shown in fig. 1, for the simple task of mapping from two input variables (a and b) to an output variable (c). (a) The prediction neurons have RFs in the two input spaces (defined by the weights $\mathbf{W}_a$ and $\mathbf{W}_b$) that make them selective to specific combinations of input stimuli. A population of pooling neurons receives input, via weights $\mathbf{V}_c$, from the prediction neurons in order to generate the output. The responses of the pooling neurons, $\mathbf{z}$, are calculated as a linear weighted sum of their input, *i.e.*, $\mathbf{z} = \mathbf{V}_c\mathbf{y}$. (b) The PC/BC-DIM network receives an additional source of input. Dealing with this extra partition of the input requires the definition of additional columns of feedforward synaptic weights, $\mathbf{W}$, and additional rows of the feedback weights, $\mathbf{V}$. If the additional feedback weights, $\mathbf{V}_c$, are identical to the pooling weights used in the architecture shown in (a), then (given eq. 1), the responses of the third partition of the reconstruction neurons, $\mathbf{r}_c$, will be identical to the responses of the pooling neurons in (a), *i.e.*, $\mathbf{r}_c = \mathbf{V}_c\mathbf{y}$. If the feedforward weights associated with the third partition, $\mathbf{W}_c$, are rescaled versions of the corresponding additional feedback weights, $\mathbf{V}_c$, then the network can perform mappings not only from a and b to c, but also from a and c to b, and from b and c to a (as will be illustrated in fig. 6).

or "dictionary" of possible representations, or more generally a model of the external environment (Spratling, 2012c, 2014a). The activation dynamics described above result in the PC/BC-DIM algorithm selecting a subset of active prediction neurons whose RFs (which correspond to basis functions) best explain the underlying causes of the sensory input. The strength of activation reflects the strength with which each basis function is required to be present in order to accurately reconstruct the input. This strength of response also reflects the probability with which that basis function (the preferred stimulus of the active prediction neuron) is believed to be present, taking into account the evidence provided by the input signal and the full range of alternative explanations encoded in the RFs of the whole population of prediction neurons.

2.3 The PC/BC-DIM Network as an Omni-directional Basis Function Network

As described in the preceding paragraph, the prediction neurons in a PC/BC-DIM network behave like basis function neurons. Figure 3 illustrates how this can be exploited to perform a simple mapping from two input variables to an output variable, analogous to the task performed by the traditional basis function network shown in fig. 1. If a sub-set of the prediction neurons represent combinations of inputs that correspond to the same value of the output, then it is necessary to "pool" the responses from this sub-set of prediction neurons to produce this output whenever one of these combinations is presented to the inputs. Figure 3 shows two ways in which this can be implemented. The first method (fig. 3a) involves using a separate population of pooling neurons that are activated by the responses of the prediction neurons. This method has been used in previous work (De Meyer and Spratling, 2013; Spratling, 2009, 2014a) and is directly equivalent to a standard basis function network. The second method (fig. 3b) involves defining additional neurons within the reconstruction neuron population that perform the same role as the pooling neurons in the first method (Muhammad and Spratling, 2015, 2017a,b; Spratling, 2016a,b). In this article the latter method will be used, as it has a major advantage over the former. Specifically, it allows omni-directional mapping between coordinate systems. For example, the network shown in fig. 3b can infer c given a and b (as illustrated in fig. 6a), but it can also infer b given a and c (as illustrated in fig. 6b), and determine a given inputs b and c (as illustrated in fig. 6c). This ability to calculate mappings in all directions is exploited in the simulations presented in this article in order to perform sensory-sensory mappings

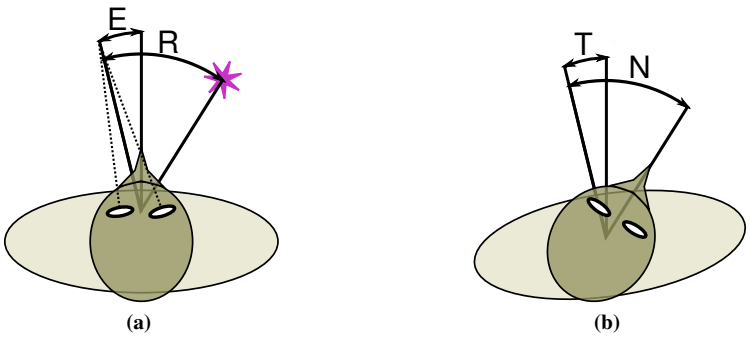


Figure 4: The simulated subject viewed from above. The agent can perform horizontal eye, neck and torso movements only. For simplicity only a single, cyclopean, eye is considered in the simulations. Furthermore, the axes of rotation of the cyclopean eye, the head and the torso are assumed to be colinear so that the simulated agent is purely linear. The anglular position of a visual stimulus on the retina is denoted by R and is measured from the centre of the cyclopean retina. The pan angle of the eye is denoted by E and is measured relative to the head. The pan angle of the neck is denoted by N and is measured relative to the torso. The pan angle of the torso is denoted by T and is measured relative to the world. (a) Shows E and R when the subject has neck and torso pan angles of zero. (b) Shows T and N when the subject has non-zero neck and torso pan values.

to determine the location of a visual target, to perform sensory-motor mappings to plan eye movements, and to perform motor-sensory mappings to predict the visual input expected following an eye movement (described in detail in sect. 2.5).

2.4 Wiring-up the Networks to Perform the Simulations

PC/BC-DIM networks of the form described above will be used to simulate gaze control. To simplify the task only horizontal gaze shifts are considered. Furthermore, only one, 1-dimensional, cyclopean eye is considered, and the simulated agent (see fig. 4) is considered to be a linear system, such that that position of an object in the world (measured in degrees from a mid-line passing through the agent's body) can be calculated as the sum of the retinal position of the target object (in degrees), the eye's orientation (in degrees, measured relative to the head), the neck's pan angle (in degrees, measured relative to the body), and the orientation of the torso (in degrees, measured relative to the world). However, it should be noted that the methods described here are not limited to such a simple, linear, case. As with traditional basis function networks, a PC/BC-DIM network can be used to approximate any linear or nonlinear mapping. Furthermore, the proposed method scales up to nonlinear systems with many more DOFs than the one considered here. Specifically, it has been used to control eye movements in a humanoid robot with two eyes, each of which produces a two-dimensional image and has two DOFs of movement (Muhammad and Spratling, 2015) and to control the humanoid robot's neck movements (three DOFs) and arm movements (three DOFs) (Muhammad and Spratling, 2017a,b).

The format used in figs. 2 and 3 to draw PC/BC-DIM networks is unwieldy, especially when illustrating larger, hierarchical, PC/BC-DIM networks. Hence, throughout the remainder of this article the simplified format, illustrated in fig. 5a, is used. The mathematical model remains unchanged, it is just the way of illustrating this model that has been simplified.

To simulate the coordinated control of eye, head and body movements the hierarchical PC/BC-DIM network illustrated in fig. 5b was used. This network consisted of a hierarchy of three processing stages. The first processing stage (shown on the left of fig. 5b) contained three partitions. The first partition represented the retinotopic input (R), the second partition represented the eye orientation (E), and the third partition represented the visual world in a head-centred coordinate system (H). This head-centred representation was invariant to eye movements.

The second processing stage also contained three partitions. The first partition represented the head-centred bearing of visual targets (H). The input to this partition came from the reconstruction neurons of the third partition of the first processing stage. The reconstruction neurons of the first partition of the second processing stage provided reciprocal input to the third partition of the first processing stage. The second partition represented neck orientation (N), and the third partition represented the visual world in a body-centred coordinate system (B). This body-centred representation was invariant to both eye and neck movements.

The third processing stage also contained three partitions. The first partition represented the body-centred

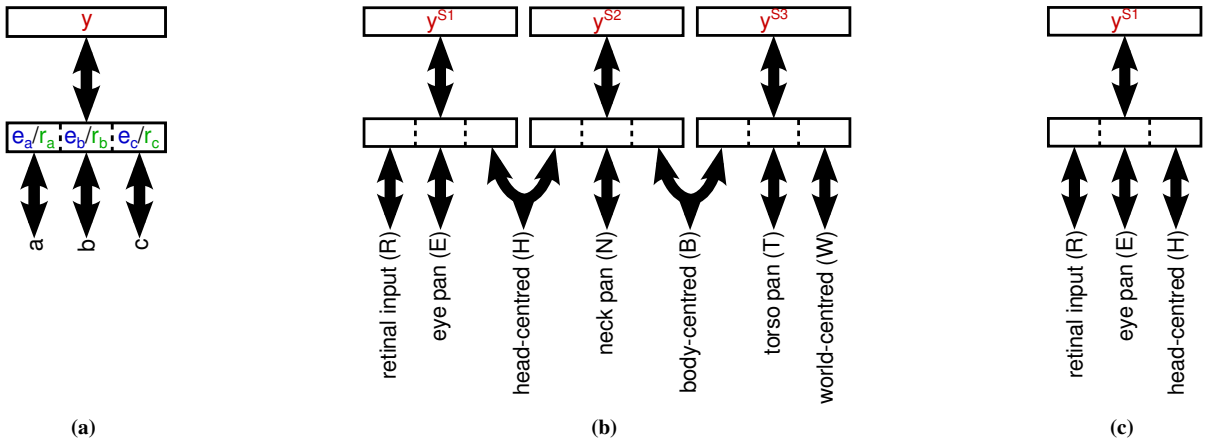


Figure 5: (a) The PC/BC-DIM network shown in fig. 3b drawn using a simplified format: the error neuron and reconstruction neuron populations are shown superimposed and double-headed arrows are used to show inputs and outputs to and from both these populations. (b-c) The PC/BC-DIM networks used to perform the simulations presented in this article, drawn using the same simplified format. These networks are used to perform (b) coordinated movements of the eye, head, and body (controlled in world-centred coordinates), and (c) the eye-movements (controlled in head-centred coordinates). The network shown in (b) was used to perform the simulations reported in sect. 3.1. The simulations described in sects. 3.2 and 3.3 were performed using the simpler network shown in (c). In (b) the curved double-headed arrows linking processing stages represent connections going from the reconstruction neurons to the input of the other processing stage.

bearing of visual targets (B). The input to this partition came from the reconstruction neurons of the third partition of the second processing stage. The reconstruction neurons of the first partition of the third processing stage provided reciprocal input to the third partition of the second processing stage. The second partition represented torso orientation (T), and the third partition represented the visual targets in a world-centred reference frame (W). This world-centred representation was invariant to eye, neck and torso orientation.

For each processing stage, each prediction neuron had a Gaussian RF (with a standard deviation of 7.5^o) in each of the three partitions. These RFs were centred in each input space so as to encode the desired relationship between the three variables for one specific set of values. The RFs of the prediction neuron population as a whole evenly tiled the input spaces, so that the network could approximate the mapping between the variables for all possible values. Hence, in the first processing stage, each prediction neuron had synaptic weights that defined a preference for one particular value of retinal input (R), one value for eye position (E), and the corresponding value of head-centred space (H). Together, a population of 187 prediction neurons were used to map between all valid combinations of R, E, and H. The prediction neuron weights in the second processing stage were defined analogously, so that each prediction neuron encoded a triplet of values satisfying the desired mapping between the three variables, H, N, and B. This second processing stage contained 513 prediction neurons. Similarly, the third processing stage contained 405 prediction neurons that acted as basis functions to implement the mapping between the variables B, T and W. [b]

It should be noted that it would also be possible to perform the simulations with a single PC/BC-DIM process-

[b]The synaptic weights of the PC/BC-DIM network have been hard-coded rather than learnt. While this article does not consider learning, it is possible to speculate as to how the weights could be learnt. Consider the first processing stage. The prediction neurons need to learn RFs that represent combinations of retinal input and eye position so as to tile this joint input space. This can be achieved using an unsupervised learning algorithm trained using randomly positioned visual targets and random eye movements, as has been shown in De Meyer and Spratling (2011). The reconstruction neurons in the third partition of the first processing stage need to learn strong connections to all the prediction neurons that represent the same head-centred location. This could potentially be achieved using an unsupervised learning rule that forms associations across time (Földiák, 1991; O'Reilly and McClelland, 1992; Spratling, 2005; Templeman and Loew, 1989; Wallis, 1996). Specifically, one reconstruction neuron could be connected to all the prediction neurons that represent a single head-centred location by training with a stationary visual target and random eye movements, and this process repeated with different target positions in order to train all the reconstruction neurons. This has been demonstrated by Spratling (2009) for a PC/BC-DIM architecture where the head-centred representation is learnt by pooling neurons, as shown in fig. 3a, rather than reconstruction neurons. Once the weights for the first processing stage have been learnt, an analogous method could be used to train the second and subsequent processing stages in a greedy layer-wise process, as is typically used to train deep neural networks (Hinton et al., 2006; Larochelle et al., 2009). Such a process might be facilitated in the human visual system by the sequential development of motor control in young infants, in which eye movements occur before the mastery of head and body movements, and head movements are performed before body movements.

ing stage that had five partitions representing the variables R, E, N, T, and W. However, as with all basis function style networks, the number of neurons required to perform transformations (with a given degree of accuracy) increases exponentially with the number of input variables (Deneve and Pouget, 2003; Pouget and Sejnowski, 1997). While this is not an issue for the simple agent simulated here, it is an issue faced by the brain in controlling a body with very many DOFs and many sources of sensory input. For such a system, using a single basis function network is unlikely to offer a tractable solution. This has almost certainly led to the brain employing a hierarchy of basis functions type neural networks (Chinellato et al., 2011; Pertzov et al., 2011; Pouget et al., 2002; Pouget and Sejnowski, 1997; Pouget and Snyder, 2000) which allow information to be transformed between multiple different coordinate systems, as is observed along the dorsal pathway of the cortical visual system (Battaglia-Mayer et al., 2003; Blangero, 2008; Colby, 1998; Marzocchi et al., 2008; McGuire and Sabes, 2009; Pertzov et al., 2011; Weber et al., 2007).

In the proposed model, each externally generated variable - retinal position (R), eye orientation (E), neck orientation (N), and torso orientation (T) - ranged over values comparable to those found in humans. Specifically, R ranged from -80^o to $+80^o$, E ranged from -50^o to $+50^o$, N ranged from -90^o to $+90^o$, and T ranged from -40^o to $+40^o$.

The inputs, $\mathbf{x}$, to each partition of each PC/BC-DIM network were encoded as Gaussian population codes, in a similar way to has been used by many related models (*e.g.*, Cassanello and Ferrera, 2007; De Meyer and Spratling, 2011; Pouget and Sejnowski, 1997; Salinas and Abbott, 1996; Zipser and Andersen, 1988). These population codes were created using a topographically-organised population of neurons with Gaussian RFs. The Gaussian RFs all had the same standard deviation (σ) but had RFs centres (μ_i) that were uniformly spaced over the range of possible values for that input variable. The response of each neuron, i, was calculated as

$$\mathbf{x}_i = \sum_p \exp\left(-\frac{(a_p - \mu_i)^2}{2\sigma^2}\right) \tag{4}$$

Where a_p is a value that needs to be encoded. In all the simulations reported here, the spacing between RF centres was 5^o and a value of $\sigma = 12.5^o$ was used.

The outputs from each partition were the responses of the reconstruction neurons, $\mathbf{r}$. Since these neurons reconstruct the input, they also generate population codes, and each neuron in the $\mathbf{r}$ population has an RF centre at the same location (μ_i) as the corresponding neuron in the input population. Decoding these values was performed using standard population vector decoding (Georgopoulos et al., 1986) to find the mean of the distribution of responses:

$$a = \frac{\sum_i r_i \mu_i}{\sum_i r_i} \tag{5}$$

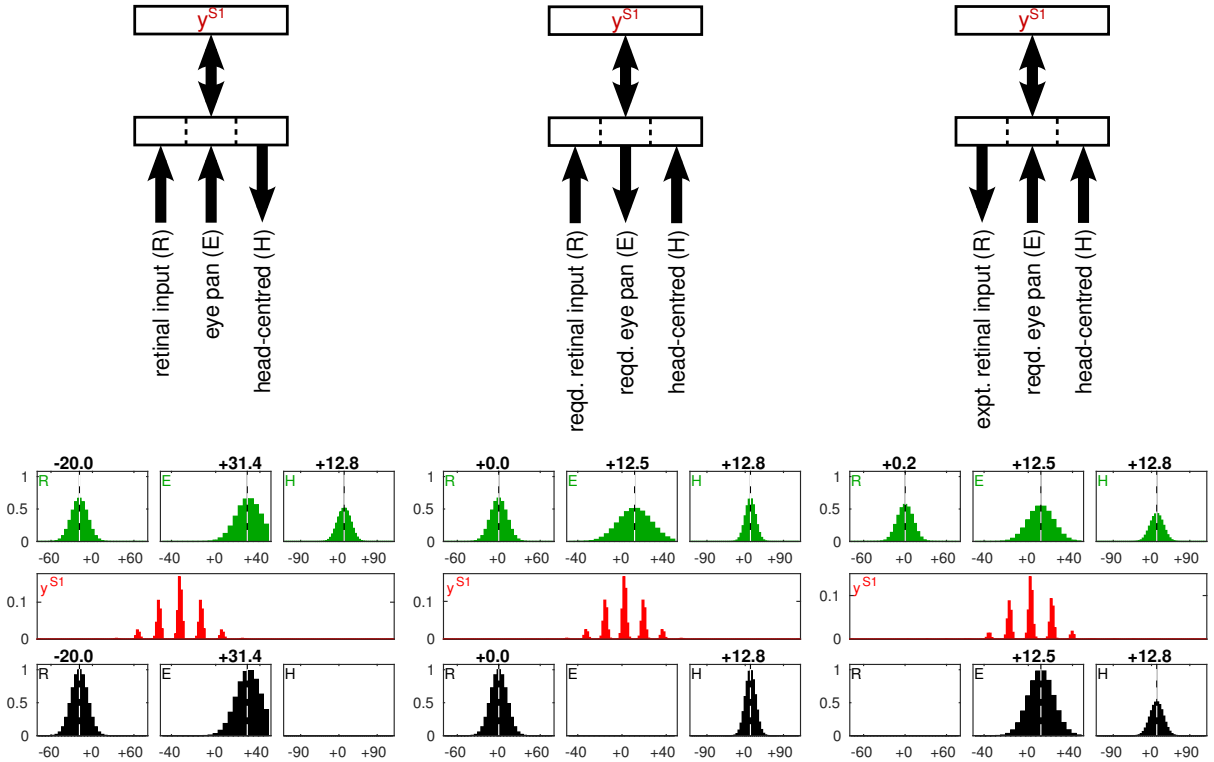
Such decoding assumes that there is only one value, a, encoded by the population code. This assumption held for the simulations performed here, except that shown in fig. 14.

Note that while the $\mathbf{x}$ and $\mathbf{r}$ populations were arranged in an orderly, topological, fashion in model, there is no requirement for them to be arranged in such an orderly map. The neurons could be physically located anywhere within each population (and neurons from different populations could be intermixed) and this would have no effect on the functioning of the model. Furthermore, it should be noted that the model uses a very impoverished representation of the visual world, consisting of typically one (but potentially several) targets represented by peaks in a population code. It is assumed that these targets correspond to the most salient locations in the visual input. The limitation to only a few such targets is consistent with the biological visual system which is limited to remembering only the three or four most salient stimulus locations across saccades (Melcher and Colby, 2008).

Many of the experiments performed in this article simulate eye movements when the head and body are fixed. Such eye movements were simulated using the simpler network shown in fig. 5c. This network is simply the first processing stage of the hierarchical network shown in fig. 5b. This simplification means that gaze shifts are planned using head-centred rather than world-centred coordinates. The experiments performed in sect. 3.1.2, with a fixed body posture, can also be performed in a body-centred coordinate system, using the first two processing stages of the hierarchical network shown in fig. 5b with little effect on the results. However, the results presented here were produced with the full network (fig. 5b), but with an input representing the fixed position of the body always applied to the second partition of the third processing stage.

2.5 Performing Gaze Control with the PC/BC-DIM Network

The procedure for performing gaze control will be described initially for the simple case where eye movements are controlled in head-centred space using the network shown in fig. 5c. To move the eye to foveate a target, three steps are performed.



(a) Step 1: mapping the visual location of the target (R) and the eye position (E) to the head-centred location of the target (H)

(b) Step 2: mapping the head-centred location of the target (H) and the required retinal location of the target (R) to the required eye position (E), *i.e.*, the planned movement

(c) Step 3: mapping the head-centred location of the target (H) and required eye position (E) to the retinal input (R) expected after the planned movement

Figure 6: The steps involved in planning gaze shifts in head-centred space. This is achieved by using the PC/BC-DIM network shown in fig. 5c to perform three operations. Each column in the figure corresponds to one operation. The upper row shows the inputs to and outputs from the PC/BC-DIM model (as illustrated in fig. 3b, inputs are presented to the error neurons, and outputs are read from the reconstruction neurons). The lower row shows the responses of the neurons in the PC/BC-DIM network for a specific example, in which the eyes are initially looking 35^o to the right and the target for the saccade is initially centred at -20^o on the retina. In the figures on the bottom row, the lower histogram shows the input to the PC/BC-DIM network (each of which is a Gaussian population code describing the value of the corresponding variable). The y-axis represents the strength of the input, and the x-axis represents the value of the variable represented by that input neuron. The middle histogram shows the responses of the prediction neurons. The y-axis is in arbitrary units representing firing rate and the x-axis is neuron number. The upper histograms show the responses of the reconstruction neurons. The y-axis represents the firing rate of the neuron, and the x-axis represents the value of the variable represented by that reconstruction neuron. The numbers above the histograms indicate the value of the variable decoded using eq. 5. (a) In the first step, the retinal position of the target and the eye position are provided as inputs to the PC/BC-DIM network. At the end of the iterative process described in sect. 2.2, the reconstruction neurons in the third partition represent the location of the target in a head-centred reference frame. (b) In the second step, the PC/BC-DIM network again receives two inputs. The first is the head-centred position of the target that was calculated in the proceeding step. The other input encodes the desired retinal position of the target following the up-coming eye movement. If the aim is to foveate the target then the retinal input is a Gaussian population code centred at 0^o,as shown here. Given these inputs, the reconstruction neurons in the second partition will, at the end of the iterative process described in sect. 2.2, represent the pan angle of the eye needed to foveate the target. (c) In the third step, both values calculated in the proceeding two steps are provided as input. The output at the end of the iterative process described in sect. 2.2, is an estimate (encoded by the reconstruction neurons in the first partition) of the visual input expected if the eye movement is performed.

The first step is to calculate the position of the visual target in head-centred coordinates. To do this, the retinal position of the target and the current eye position are provided as inputs to the PC/BC-DIM network, as illustrated in fig. 6a. The two input distributions cause responses in the subset of prediction neurons with RFs that are centred near the visual input in the first partition and near the eye pan input in the second partition. Each of these prediction neurons has an RF centred near the corresponding head-centred location of the target in the third partition (due to wiring of the network described in sect. 2.4). The reconstruction neuron responses are a linear combination of all the active prediction neuron RFs, and hence, will peak at the appropriate places in each of the three partitions. Hence, the response of the reconstruction neurons in the third partition provides a representation of the head-centred location of the target, which can be read out from the network.

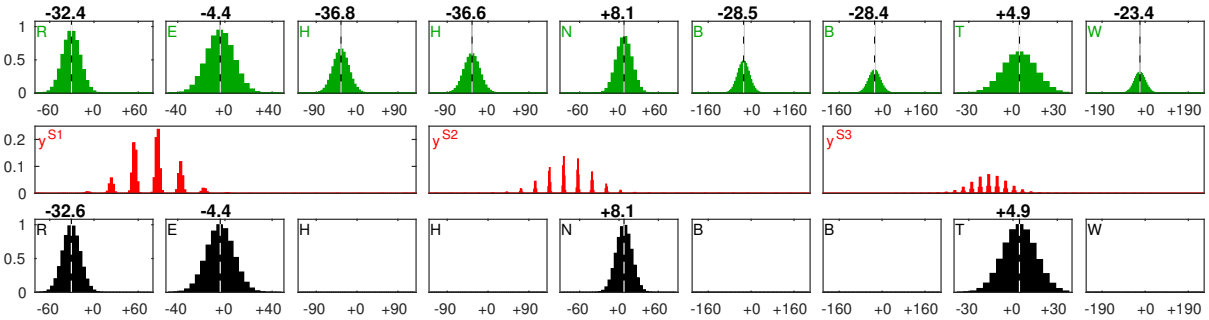
The second step is to calculate the eye position required to foveate the target. To do this, the required retinal position of the target and the head-centred position of the target calculated in the first step are provided as inputs to the PC/BC-DIM network, as illustrated in fig. 6b. Given these inputs, the PC/BC-DIM algorithm will reach a steady-state in which prediction neurons (basis function neurons) that best represent the inputs will be selected to be active. Due to wiring of the network (see sect. 2.4), these neurons will also represent the eye orientation required foveate the target, and hence, the required eye position can be read out from the reconstruction neuron responses in the second partition.

The third step is to calculate the retinal position of the target expected following the forthcoming eye movement. To do this, both the values calculated in the preceding two steps (the head-centred location of the target and the planned eye position after the gaze shift) are provided as inputs to the PC/BC-DIM network, as illustrated in fig. 6c. Performing the iterative process described in sect. 2.2 will result in the reconstruction neurons representing the estimated visual input expected after the saccade to the target.

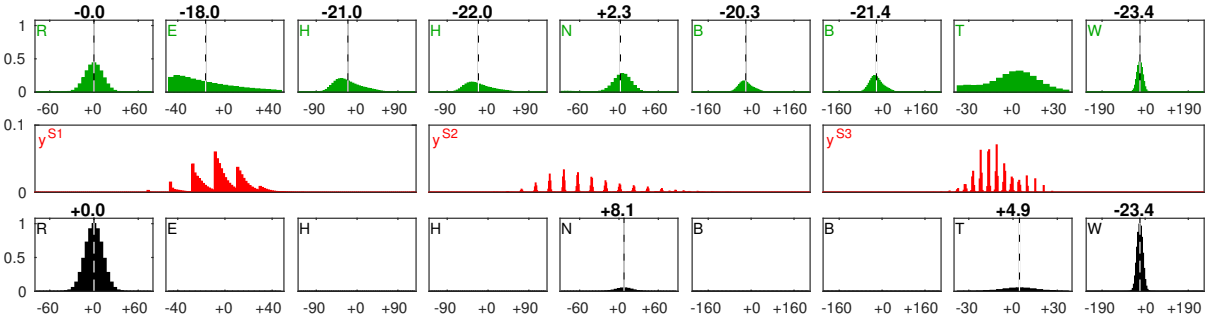
The first step performs a sensory-sensory mapping to determine the location of a visual target, the second step performs a sensory-motor mapping to plan the eye movement, and the third step performs a motor-sensory mapping to predict the visual input expected after following an eye movement. This latter step is not essential for eye movement planning, but is responsible for RF remapping (as will be discussed in sect. 3.3.2). Typically, following the third step, the simulated agent performs the planned eye movement. These steps may be repeated to perform corrective saccades if the initial eye movement fails to accurately bring the target onto the fovea. Such corrective saccades have also been observed in head-restrained human subjects (Cohen and Ross, 1978; Henson, 1978; Kapoula and Robinson, 1986). At the start of the second step the prediction neuron activities are set to zero, but for other steps the initial state of the network is given by the prediction neuron responses resulting from the preceding step. Each step is performed using 100 iterations of eqs. 1–3.

In many circumstances saccades are not produced as an immediate response to the presence of a visual target. They may be initiated through other modalities (such as sound), or may be made to a location stored in memory. To simulate delayed saccades it is assumed that having performed the above three steps the required eye pan value is placed in memory, rather than being executed immediately. During the delay period, only the first and third steps are performed in a continuous loop, using the current eye pan value rather than the planned, post-saccadic, one. At the time when the saccade is performed the required eye position is injected at the beginning of the remapping step, just prior to the memory-driven saccade being performed.

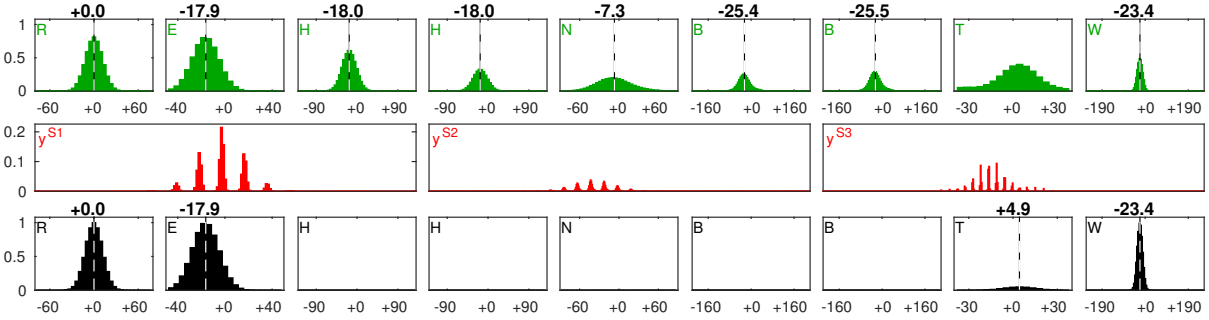
To control body, head, and eye movements using the complete network shown in fig. 5b, a similar procedure is used, except that there are two additional steps in order to calculate the required movement of the neck and torso in addition to calculating the required orientation of the eye. In effect, the second step described above is split into three steps (2.1, 2.2, and 2.3) to calculate the required orientation of each of the three DOFs. A particular example of this process is illustrated in fig. 7. The sensory-motor mapping steps determine the required orientation of each DOF in the following order: eye position (step 2.1), neck position (step 2.2), and torso position (step 2.3). At the start of step 2.1 there are only two knowns and three unknowns. The knowns are the desired retinal position of the target after the saccade, and the world-centred position of the target. The three unknowns are the eye, neck and torso orientations. To help the PC/BC-DIM network find a unique solution for the required eye orientation, weak inputs representing the current neck and torso orientations are also presented to the network (the amplitude of these weakened inputs is controlled by parameter ψ which is set to 0.05 in all experiments). By using the current values for neck and torso position, this induces a kind of "inertia" in these DOFs, and subsequently biases the network to prefer making eye movements over head and body movements. The steps described above to plan body, head, and eye movements may be repeated to perform corrective gaze shifts as observed in primates (McCluskey and Cullen, 2007). For simplicity, the current model performs all motor actions, simultaneously, at the end of the third step. However, it would also be possible for movements to be initiated as soon as they were planned, meaning that eye movements would start prior to head movements, which would start prior to body movements. This would be consistent with the sequential onset of these different movements observed in the biological system (McCluskey and Cullen, 2007; Tweed et al., 1995).



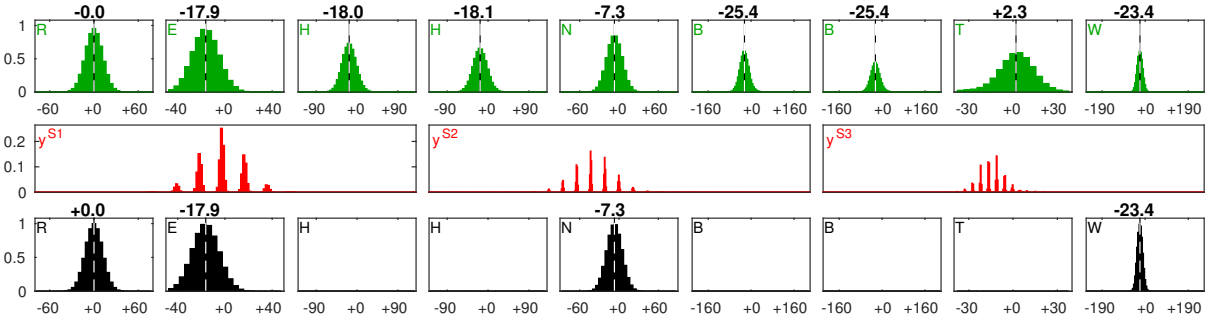
(a) Step 1: mapping the visual location of the target (R), and postural inputs (eye (E), neck (N) and torso (T) positions) to the world-centred location of the target (W)



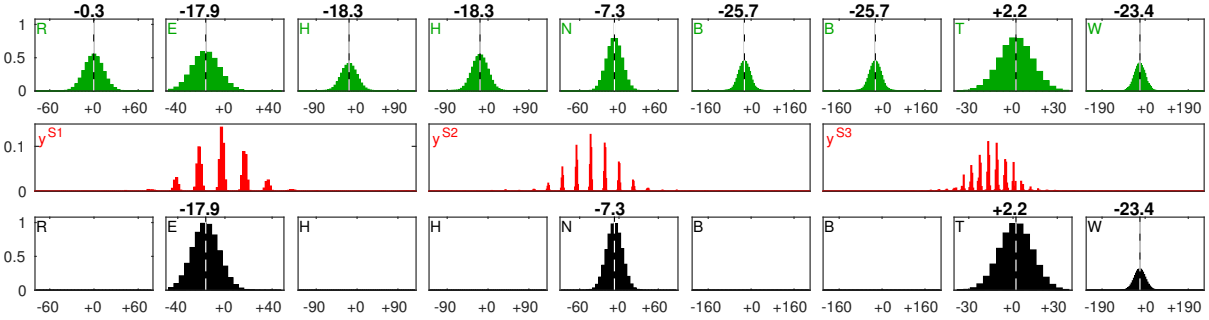
(b) Step 2.1: mapping the world-centred location of the target (W) and the required retinal location of the target (R) to the required eye position (E) *i.e.*, the planned eye movement



(c) Step 2.2: mapping the world-centred location of the target (W), the required retinal location of the target (R), and the required eye position (E) to the required neck position (N) *i.e.*, the planned neck movement



(d) Step 2.3: mapping the world-centred location of the target (W), the required retinal location of the target (R), the required eye position (E), and the required neck position (N) to the required torso position (T) *i.e.*, the planned body movement



(e) Step 3: mapping the world-centred location of the target (W) and required eye (E), neck (N) and torso (T) positions to the retinal input (R) expected after the planned movement

Figure 7: (previous page) The steps involved in planning gaze shifts in world-centred coordinates. This is achieved by using the PC/BC-DIM network shown in fig. 5b. Each row in the figure corresponds to one operation, and has the same format as the lower row in fig. 6. Each row shows the responses of the neurons in the PC/BC-DIM network for a specific example of gaze control in which the eyes are initially oriented at -4.4^o, the neck orientation is $+8.1^o$, the initial torso orientation is $+5^o$, and the target to be foveated appears on the retina at -32.6^o. (a) Step 1: calculates the world-centred representation of the visual target (W) given the retinal position of the target (R), the eye position (E), the neck position (N), and the torso position (T). (b) step 2.1: calculates the eye position (E) required to foveate the target. The inputs are the world-centred position (W) calculated in step 1, weakened inputs representing the current neck (N) and torso positions (T) and the desired retinal input (R), centred at the fovea. (c) Step 2.2: calculates the neck position (N) required to foveate the target. The inputs are the world-centred position (W) calculated in step 1, the eye position (E) calculated in step 2.1, a weak input representing the current torso position (T), and the desired retinal input (R), centred at the fovea. (d) Step 2.3: calculates the torso position (T) required to foveate the target. The inputs are the world-centred position (W) calculated in step 1, the eye position (E) calculated in step 2.1, the neck position (N) calculated in step 2.2, and the desired retinal input (R), centred at the fovea. (e) Step 3: calculates the visual input that is expected after the upcoming saccade given the world-centred representation (W) and the planned eye (E), neck (N) and torso (T) positions.

2.6 Code

Open-source software, written in MATLAB, which performs the experiments described in this article is available from: `http://www.corinet.org/mike/Code/pcbc_gaze.zip`.

3 Results

3.1 Coordinated Eye, Head and Body Movements

Typically, movements that change the direction of gaze can be performed using movements of the eyes, head and body. To perform an accurate movement, one that brings the target location onto the fovea, the brain is faced with two challenges: finding *valid* and *unique* positions in which to place each DOF. Finding a valid solution requires coordination between the planned movements of the different DOFs: it is no good independently planning movements of the eye and the neck if the combined movement fails to foveate the target, nor is it useful to plan torso and neck postures that would place the target beyond the physical range of movement that is possible for the eye. There may be multiple valid solutions for the same target location due to the visual system being highly redundant. This means that, typically, there are many different combinations for eye, neck, and torso orientations that could be used to foveate the same target location. The brain must choose one, unique, solution that will actually be performed. Far from being chosen at random, empirical studies have found that there are predictable relationships between the movements made by the body, head and eyes (Freedman and Sparks, 1997; Guitton and Volle, 1987; McCluskey and Cullen, 2007; Tomlinson and Bahra, 1986).

The following simulations demonstrate that the PC/BC-DIM model produces movements with similar relationships between body, head and eye movements. These simulations of coordinated eye, head and body movements are performed using the PC/BC-DIM model shown in fig. 5b using the procedure illustrated in fig. 7. The success of the model in reproducing the biological data relies on the way in which it resolves the kinematic redundancies to chose a particular combination of movements to perform each task. The population codes that represent the sensory-motor variables input to, and output by the reconstruction neurons of, the PC/BC-DIM network can be considered to be probability distributions encoding the uncertainty about the value of each variable (Spratling, 2016a). The PC/BC-DIM network can then be considered to compute functions of variables whose values are defined in probabilistic terms (Spratling, 2016a). Decoding a particular partition of the reconstruction neurons finds the maximum likelihood estimate of the value represented by that population of reconstruction neurons. Hence, decoding selects the action that is most likely to achieve the required task, given the uncertainty about the values of the inputs to the network. In contrast, most other models of motor control resolve kinematic redundancies by choosing the combination of movements that minimise a cost function (Saeb et al., 2011; Todorov, 2004), where the cost might be defined in terms of movement error, movement duration (Harris and Wolpert, 2006), energy expenditure (Kardamakis and Moschovakis, 2009), the variance of the final position in the presence of signal dependent noise (Harris and Wolpert, 1998), or changes of acceleration or of torque (Todorov, 2004).

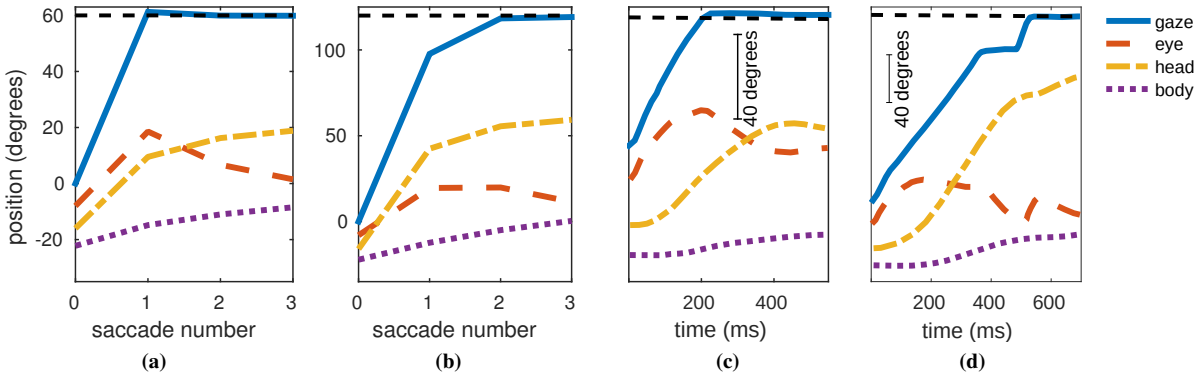


Figure 8: Examples of the movements made to perform gaze shifts of (a) 60^o, and (b) 120^o with the PC/BC-DIM model. Each graph shows (on the y-axis) the orientation of the eye, neck, and torso prior to the gaze shift (at 0 on the x-axis), after the initial gaze shift (at 1 on the x-axis) and following the first and second corrective saccade (at 2 and 3 on the x-axis). Similar results for a monkey when performing gaze shifts of (c) 60^o, and (d) 120^o (adapted from Fig. 2 of McCluskey and Cullen, 2007).

3.1.1 Eye, Head and Body Control

The control of eye, head and body movements during gaze shifts were investigated, in Macaca mulatta monkeys, by McCluskey and Cullen (2007). In these experiments, gaze shifts were induced by presenting a sequence of visual targets. The initial target was straight ahead, and subsequent targets appeared at positions which alternated between being to the left or to the right of centre, but were otherwise at unpredictable locations. The left and right targets were at eccentricities of between 20^o and 80^o in steps of 10^o. For large gaze shifts, the location of the target would not be visible to the animal, as it would appear at a position beyond the edge of the field of view. However, such gaze shifts were successfully performed, presumably as the animal could predict that the next target was beyond the current visual field in the contra-lateral half of the world. To simulate such large gaze shifts with the model, when no target was visible the world-centred target for the saccade was set to be at 50^o on the opposite side to the current gaze direction. The exact value used for these memory-driven gaze shifts had little effect on the results.

In the simulations, as in the corresponding experiments with monkeys (McCluskey and Cullen, 2007), the magnitude of the eye movement would reach a peak after which the eyes would move backwards, in the opposite direction to the gaze shift, while the neck and torso would continue to move in the direction of the gaze shift (see fig. 8). To enable the presentation and comparison of the results from many experiments the movements made on each individual trial were summarised using a number of measures. The "eye amplitude" was defined as the distance travelled by the eye before any backwards movement (McCluskey and Cullen, 2007). The "head contribution" and "body contribution" were defined as the displacement of the neck and torso at the time when the eye displacement was at its maximum (Freedman and Sparks, 1997; McCluskey and Cullen, 2007). The "head amplitude" and "body amplitude" were defined as the maximum displacement of the neck and torso (Freedman and Sparks, 1997; McCluskey and Cullen, 2007). The same measures used to quantify the psychophysical experiments were also used to summarise the results of the simulations performed here.

The results from the simulations of the experiments performed by McCluskey and Cullen (2007) are shown in fig. 9a. These simulation results are consistent with the monkey data. Specifically, for the smallest gaze shifts the eye makes the largest contribution while the body makes the least contribution. For larger gaze shifts, the eye movement saturates at well below the physical limits of eye rotation (*i.e.*, the functional range is less than the mechanical range). In contrast, the amplitude of the head and body movements continue to increase with increasing gaze amplitude. This increase is approximately linear for the head. This results in the head making the largest contribution to the overall movement for medium and large gaze shifts. In contrast, the body movement is always relatively small, even at the largest gaze shifts. Experiments with subjects who are prevented from making body movements (*e.g.*, Freedman and Sparks, 1997; Guitton and Volle, 1987; Tomlinson and Bahra, 1986, see also sect. 3.1.2) have also found that eye movements saturate below the limits of eye rotation, and that for small gaze shifts eye movements make the highest contribution while for larger gaze shifts the relative contributions of head and eye are reversed.

The two features of the model that are primarily responsible for reproducing these behaviours are firstly edge-effects with the population coding and decoding (see sect. 2.4), and secondly the "inertia" in the movement of the

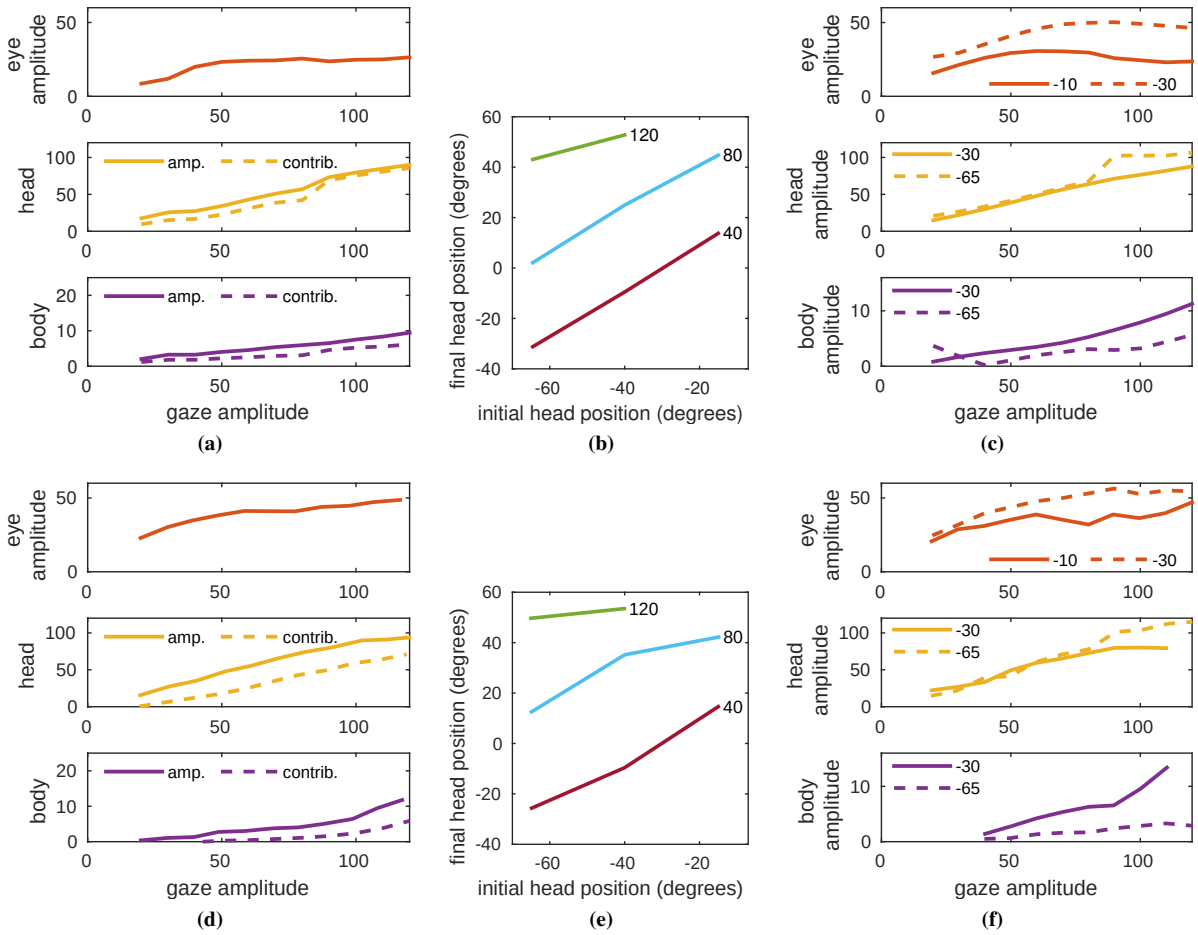


Figure 9: Simulations of coordinated eye, head and body control. (a) Relative contributions of eye, head and body movements as a function of the amplitude of the gaze shift. (d) Similar results for monkey (adapted from Fig. 5 of McCluskey and Cullen, 2007). (b) The initial and final head position for gaze shifts of amplitude 40^o, 80^o and 120^o. (e) Similar results for monkey (adapted from Fig. 6a of McCluskey and Cullen, 2007). (c) Effects of initial conditions on eye, head and body movements. The top row shows eye amplitude as a function of the amplitude of the gaze shift in trials when the initial torso position was central and the initial eye position was either 10^o or 30^o contralateral to the target. The second and third rows show the head and body movements as a function of the amplitude of the gaze shift in those trials when the initial torso position was central, the initial eye position was central, and the initial neck orientation was either 30^o or 65^o contralateral to the target. (f) Similar results for monkey (adapted from Fig. 8 of McCluskey and Cullen, 2007).

neck and torso (see sect. 2.5). Using a Gaussian population code to represent a value near the centre of the range of possible values produces accurate results. For example, in fig. 6a the value of R encoded by the first partition of the network is accurately represented as -20^o. In contrast, encoding a variable value nearer the extremes of the range of possible values produces inaccurate results due to neurons representing values beyond the edge of the range not existing. For example, in fig. 6a the true value of E is 35^o but the value encoded by the second partition of the network is 31.4^o. This bias to represent values near to the edge as being nearer to the centre of the range results in the saturation of the eye movements before they reach the extreme of the possible range of eye movements. When the network is being used to plan the eye movement, weakened inputs representing the current head and body positions are provided as input to the network. This induces inertia in the movement of the neck and torso, and hence, a preference for making larger eye movements. However, when larger gaze shifts need to be performed, the saturation of the eye movement results in larger neck and torso movements being planned in order to produce accurate foveation of the target. The relative contribution of body, head and eye are affected by the strength of parameter ψ which determines the amplitude of the weakened inputs representing the current head and body positions, and hence, the strength of the inertia. For a particular amplitude of gaze shift, higher values of

ψ lead to smaller head and body movements, while smaller values of ψ result in larger contributions by the head and body. This can be used to account for some of the individual differences observed between different animals in the monkey experiments (McCluskey and Cullen, 2007).

Just as the eye rotation saturates at about 20^o less than the physical range of possible eye rotations, so the neck movement produced by the PC/BC-DIM model will also saturate at about 20^o less than the range of possible neck rotations due to edge-effects in the population coding of neck position. McCluskey and Cullen (2007) also found that head rotations saturate before the physical limits in monkeys. They did this by recording initial and final head orientation for gaze shifts of different amplitudes when the eyes were approximately centred in their orbits. They found that the final head orientation was never greater than 60^o. The PC/BC-DIM model simulation of these results is shown in fig. 9b.

To explore the effects of initial eye position on subsequent eye movements, McCluskey and Cullen (2007) recorded gaze shifts that were made when the eye prior to the gaze shift was rotated either approximately 10^o or 30^o contra to the direction of the gaze shift. It was found that the more the eye was initially rotated away from the upcoming target, the larger the amplitude of the eye movement. The same results are obtained from the PC/BC-DIM model, as shown in the top row of fig. 9c. This is because an initial rotation of the eye away from the target means that a larger eye movement can be made before the planned eye position reaches the end of its functional range.

Similarly, to explore the effects of initial head position on subsequent head and body movements, McCluskey and Cullen (2007) recorded gaze shifts that were made when the neck prior to the gaze shift was rotated either approximately 30^o or 65^o contra to the direction of the gaze shift. It was found that the head being initially rotated further away from the upcoming target resulted in larger head movements, but only for large gaze shifts of greater than approximately 90^o. The same results are obtained from the PC/BC-DIM model, as shown in the middle row of fig. 9c. McCluskey and Cullen (2007) also found that the head being initially rotated further away from the upcoming target resulted in smaller body movements. The results from the model, as shown on the bottom row of fig. 9c, are also consistent with this data.

3.1.2 Eye and Head Control

Data obtained when subjects are restrained from making body movements, show a similar pattern of results. For example, Freedman and Sparks (1997) investigated the coordination of eye and head movements in Macaca mulatta monkeys. In these experiments, gaze shifts were induced by presenting a sequence of fixation targets at random locations on a screen, but such that all targets were within $\pm 45^o$ of the centre. Only those experiments that explored horizontal gaze shifts are simulated here. To simulate a fixed body posture, an input representing a torso orientation of 0^o was presented to the PC/BC-DIM network throughout all the steps involved in performing a gaze shift (see sect. 2.5).

The simulation results, as shown in fig. 10, are consistent with the monkey data. Specifically, for the smallest gaze shifts the eye makes a larger contribution than the head. Whereas, for larger gaze shifts, head movements contribute more to the overall gaze shift. For larger gaze shifts there is an approximately linear increase in head movement with gaze amplitude. In contrast, for larger gaze shifts the eye movement saturates at well below the physical limits of eye rotation. As the initial orientation of the eye is rotated more contralaterally to the direction of the gaze shift, so the amplitude and contribution of the head movement is reduced, and the amplitude of the eye movement is increased.

The relationship between the initial orientation of the eye and subsequent head and eye movements is made more explicit in fig. 11. These simulation results are also in close agreement with the monkey data. Specifically, head movements (measured using amplitude and contribution) are always larger for gaze shifts of greater amplitude. The head contribution for the gaze shift 25^o is close to zero for all initial eye positions, whereas the eye amplitude for the gaze shift 25^o is close to 25^o for all initial eye positions. When the initial eye position is close to zero or is ipsilateral to the direction of the gaze shift, eye amplitude is similar for all gaze amplitudes. As discussed in the previous section, the model successfully simulates this data due to edge-effects with the population coding and the "inertia" in the movement of the neck.

3.2 Saccadic Eye Movements

Saccades are ballistic eye movements performed to bring a new visual target onto the fovea (Findlay and Walker, 2012). Such eye movements are typically studied using subjects who are restrained from making head movements (*e.g.*, Albano and Wurtz, 1982; Becker and Jürgens, 1979; Goldberg and Bruce, 1990; Hallett and Lightstone, 1976; Heide et al., 1995; Honda, 1991; Komoda et al., 1973; Kusunoki and Goldberg, 2003; Mays and Sparks, 1980; Nakamura and Colby, 2002; Wetter and Opstal, 2008). Even when the subject's head is not constrained

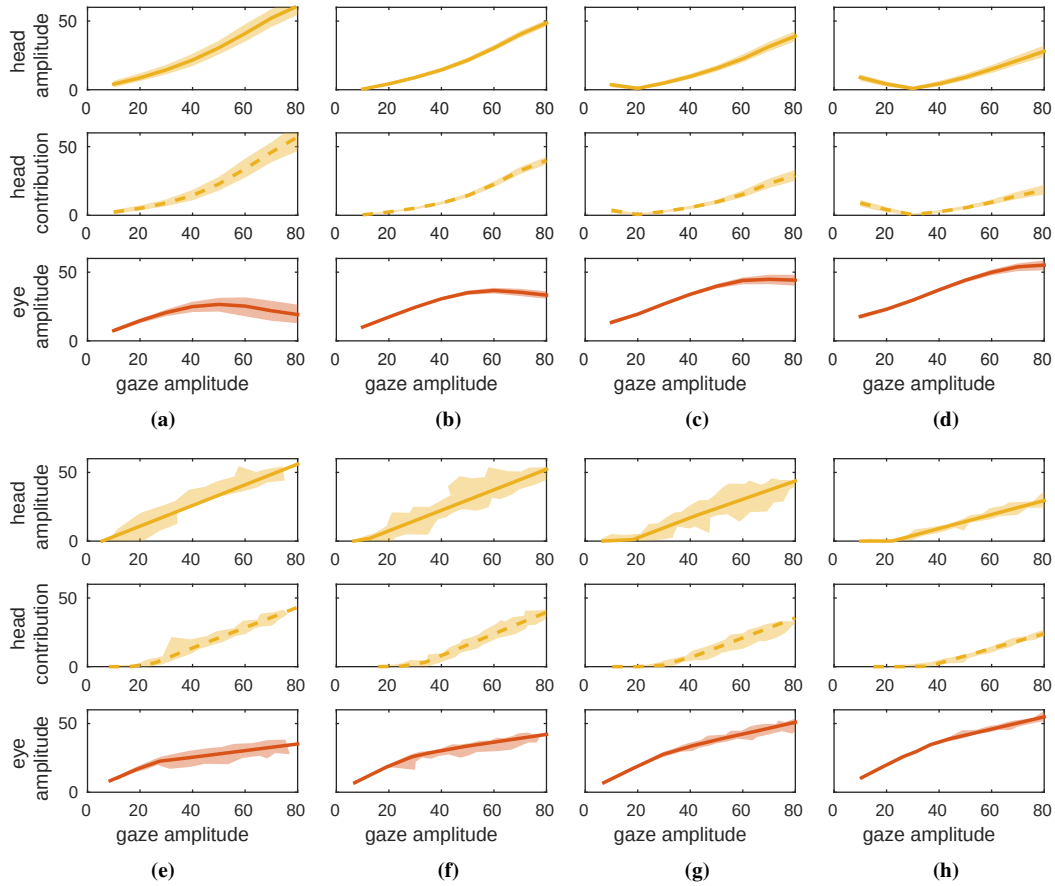


Figure 10: Simulations of coordinated eye and head control. Each figure shows eye and head movements as a function of the amplitude of the gaze shift. The initial rotation of the eye prior to the gaze shift was approximately (a) 0^o, (b) 10^o, (c) 20^o, or (d) 30^o contra-lateral to the direction of the gaze shift. Solid lines show the mean, and the shaded area show the range of results obtained under different initial conditions. Similar results for monkey when the initial rotation of the eye prior to the gaze shift was approximately (e) 0^o, (f) 10^o, (g) 20^o, or (h) 30^o contra-lateral to the direction of the gaze shift (adapted from Fig. 14 of Freedman and Sparks, 1997).

only eye movements are analysed (*e.g.*, Lappe et al., 2000; Michels and Lappe, 2004). Simulations of saccadic eye movements are performed using the PC/BC-DIM model shown in fig. 5c using the procedure illustrated in fig. 6.

3.2.1 Accuracy of Saccades

In primates, saccades to locations within the central 20^o of the visual field are accurate to within $1-2^o$ (Albano and Wurtz, 1982). Beyond the central region, or when the amplitude of the saccade is large, the saccade typically undershoots the target location (Albano and Wurtz, 1982; Leigh and Kennard, 2004). To investigate saccadic accuracy in the PC/BC-DIM model, eye movements were performed to visual targets at different eccentricities from an initial fixation point at 0^o. The results, shown in fig. 12a, are consistent with the biological data. Specifically, saccades made to targets within 20^o of fixation were accurate to within 0.8^o, while targets at greater eccentricities were undershot. This undershoot is due to the population coding edge-effects that also contributed to the saturation of eye movements at less than physical limit which was observed in sect. 3.1.

Albano and Wurtz (1982) tested saccade accuracy in Macaca mulatta monkeys. The monkeys were required to make fixed amplitude saccades of 10^o from different initial positions. The direction of the saccades was either towards or away from 0^o. The results from simulations of these experiments are shown in figs. 12b and 12c. The inaccurate foveation of both the start and end positions for saccades performed in the periphery are similar to, but less extreme than, the results obtained by (Albano and Wurtz, 1982) after ablation of the superior colliculus and neighbouring sub-cortical structures. This is to be expected, as the model described here is a model of cortex and

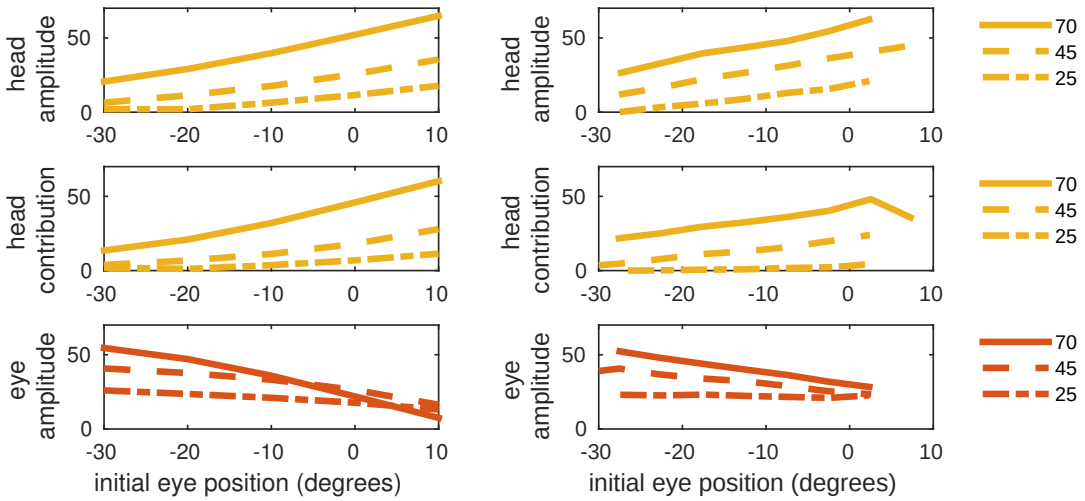


Figure 11: Simulations of coordinated eye and head control. (a) Eye and head movements as a function of initial eye position. Results are shown for gaze shifts of 25^o, 45^o, and 70^o. (b) Similar results for monkey (adapted from Fig. 15 of Freedman and Sparks, 1997).

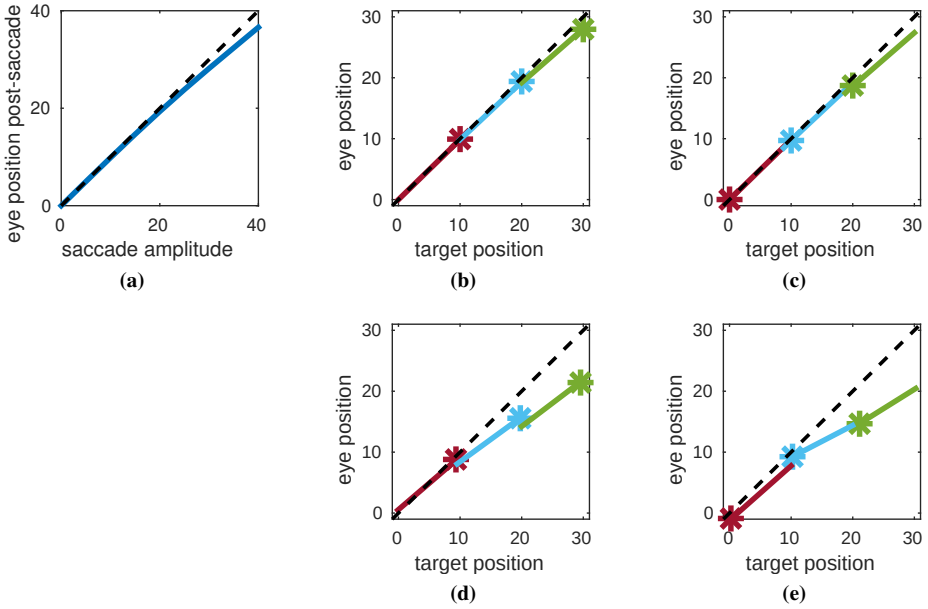


Figure 12: Accuracy of Saccades. (a) Accuracy of saccades of variable amplitude made from a central eye position. The x-axis shows the saccade amplitude, which is equivalent to the position of the target. The y-axis shows the eye position after the saccade. The plot would follow the dashed diagonal line if all saccades were perfectly accurate. (b) and (c) Accuracy of saccades of fixed amplitude (10^o) made from different starting eye positions. The end position of each saccade is indicated by a star. The direction of the saccade is towards more peripheral targets in (b) and towards more central targets in (c). (d) and (e) Similar results for monkey after ablation of the superior colliculus and neighbouring sub-cortical structures (adapted from Fig. 6 and 7 of Albano and Wurtz, 1982).

does not simulate sub-cortical mechanisms. The saccadic inaccuracies in the ablated monkeys were not related to either the amplitude or direction of the saccade, but to the orbital eccentricity of the targets (Albano and Wurtz, 1982). This is also the case for the PC/BC-DIM model, and is consistent with the cause being edge-effects in the population code representing eye pan.

3.2.2 Memory-Guided and Double-Step Saccades

Adult primates are capable of performing a sequence of saccades to different visual targets. This is the case even if the target for the next saccade is no longer visible due, for example, to it being placed beyond the edge of the visual field by the preceding saccade, or due to occlusion by a moving object. Performing such eye movements requires the subject to maintain a memory of the location of the previously visible target. Such memory-guided saccades are studied in the laboratory using the double-step saccade task (Becker and Jürgens, 1979; Hallett and Lightstone, 1976; Heide et al., 1995; Komoda et al., 1973; Mays and Sparks, 1980; Westheimer, 1954). Typically, in these experiments a visual target jumps from the initial fixation spot (location zero) to another position (location one), then makes a second jump to location two, before either being extinguished or returning to the starting position. The subject is required to follow this sequence of movements with their eyes. A particularly interesting condition occurs when the sequence followed by the visual target has been completed before the subject's eyes begin to move (Heide et al., 1995; Mays and Sparks, 1980). In this case, the subject needs to make a memory-guided saccade to both locations. Furthermore, the planned movement to foveate the second location must take into account that this movement will be performed from location one, and not from location zero (where the eyes were focused when the target was seen).

One possibility is that the second location is encoded in a retinocentric representation, and that this representation is updated to take account of the first saccade (Duhamel et al., 1992; Goldberg and Bruce, 1990). An alternative theory is that the initial retinal position of the second target, together with the initial eye position, is used to create a representation (such as a head-centered representation) of the target location that is invariant to eye movements, and for the second saccade to be planned using this invariant representation (Mays and Sparks, 1980; Robinson, 1975). The PC/BC-DIM algorithm can be used to implement a model of the second type. Specifically, the network shown in fig. 5c could be used to calculate the head-centred location of the first location (by following the procedure illustrated in fig. 6a). This representation could be stored in memory. When the target moves to the second location, the same procedure could be applied to calculate the head-centred location of the second location, which could also be kept in memory. To perform the eye movements the network would initially be supplied with the head-centred location of the first target, and using the procedure illustrated in fig. 6b, determine the eye position required to foveate that location. Having performed this movement, the network could then calculate the eye position required to foveate the second location, by following the same procedure using the head-centred location (retrieved from memory) of the second target.

An alternative possibility is that what is memorised is the eye pan values required to foveate the two locations, rather than the head-centred representations of those locations. To do this the PC/BC-DIM network shown in fig. 5c could perform the first and second steps illustrated in fig. 6 in order to calculate the eye pan required to foveate the first target. This representation could be stored in memory. When the target moves to the second location, the same procedure could be applied to calculate the eye position required to look at the second location, which could also be kept in memory. Performing the saccades would then just require these two remembered eye pan values to be used to drive the eye movements.

To facilitate building such a model, PC/BC-DIM can also be used to provide a working memory, as illustrated in fig. 13. In a PC/BC-DIM network, if an input is connected to many prediction neurons then its influence is modulatory rather than driving (Spratling, 2014b). In the network presented in fig. 13 the one input in the second partition connects to all the prediction neurons and is thus modulatory. When presented in isolation this input causes only a uniform and very weak response from all the prediction neurons it connects to. However, when presented together with inputs to the first partition, which lead to the strong activation of a sub-population of prediction neurons, the second partition input can enhance the response of this sub-population and maintain its response even when the driving input to the first partition is removed. The response of the sub-population of prediction neurons will be maintained until the second partition input is removed, or until a new pattern of inputs is presented to the first partition. The second partition input will maintain activity caused by any pattern of activity in the first partition, so it acts as a general signal to store (when active) or reset (when inactive) the memory.

A completely functional model of memory-guided saccades would, however, require mechanisms additional to those described above. The additional mechanisms are required to act as a form of executive system to decide when to perform different steps shown in fig. 6 and when to transfer information to and from memory. It is possible to simply repeatedly perform all three steps shown in fig. 6, but then it is still necessary for additional, executive, functions to decide when to move information to and from the memory and when to perform the actions that are

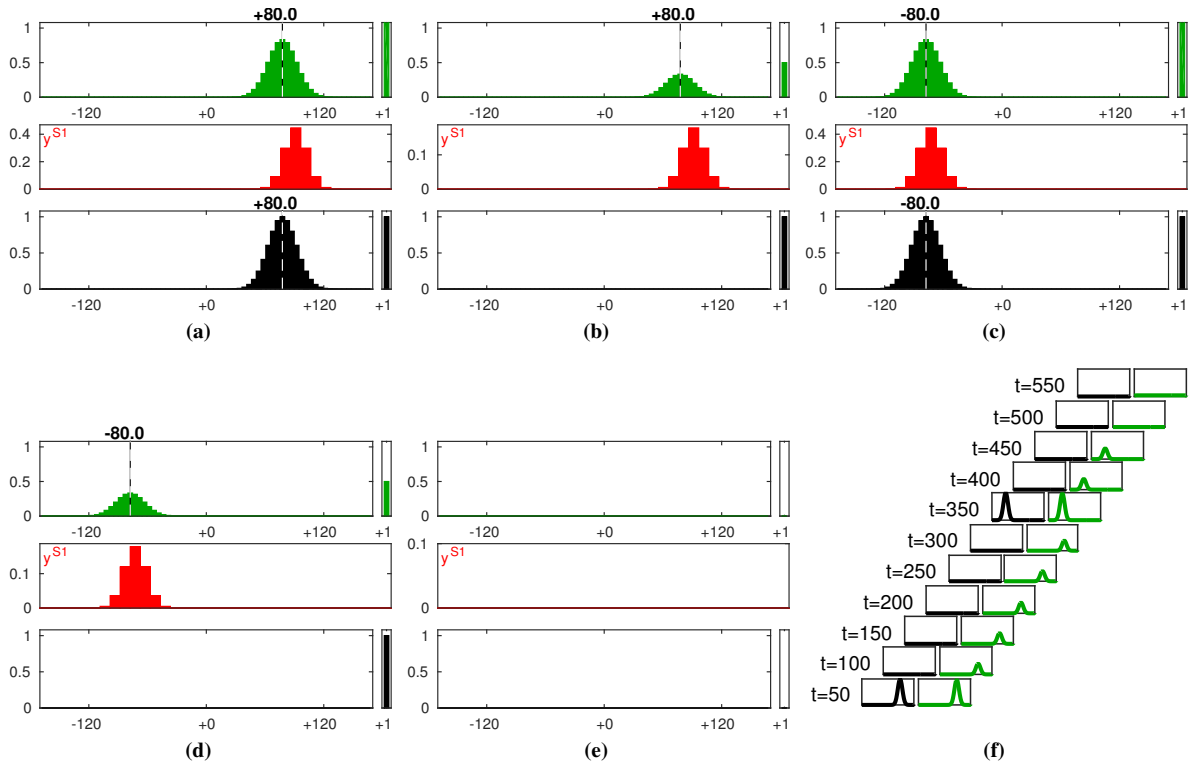
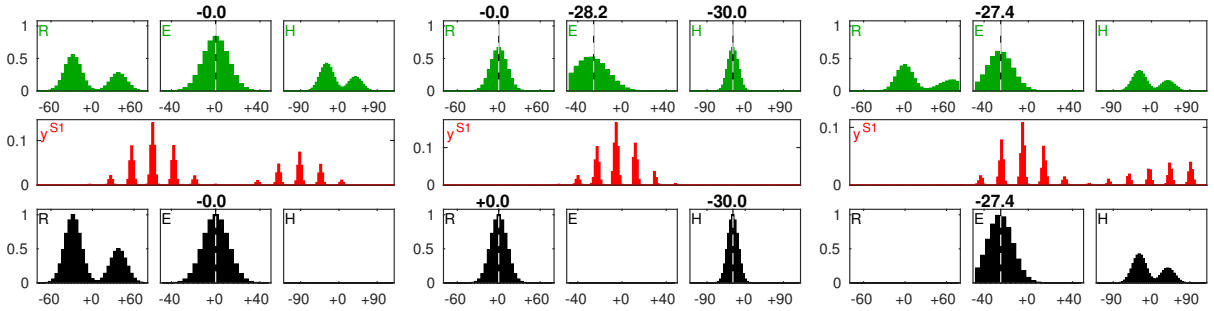


Figure 13: Implementing working memory in a PC/BC-DIM network. The network receives input from two partitions. The first represents population coded information. The second partition contains only one input, which determines whether or not information is retained in working memory. The format of the figures (a) to (e) is identical to the lower row of fig. 6, and is described in the caption of that figure. (a) The state of the network after 50 iterations during which time input has been presented to the first partition. The input to the second partition is also active. (b) The state of the network after a further 250 iterations during which time input has only been applied to the second partition. The reconstruction neurons retain a memory of the preceding input to the first partition. (c) The state of the network after a further 50 iterations during which time a new input has been presented to the first partition. The input to the second partition is also active. The new input has overwritten the memory of the first input. (d) The state of the network after a further 100 iterations during which time input has only been applied to the second partition. The reconstruction neurons retain a memory of the preceding input to the first partition. (e) The state of the network after a further 50 iterations during which time no input is applied to either the first or second partition. The absence of the input to the second partition causes the previous inputs to be forgotten. (f) The results of the same simulation showing (at 50 iteration intervals) the input to the first partition (on the left), and the corresponding reconstruction neuron responses (on the right). The results shown at t=550 iterations is an additional result, showing the reconstruction neuron responses after input has been applied to the second partition of the network for 50 iterations in the absence of any input to the first partition.

planned using the PC/BC-DIM network.

In typical laboratory-based studies on double-step saccades the two visual targets are presented in sequence. In more ecologically-valid circumstances, multiple targets for future saccades are present simultaneously. In contrast to previous basis function type networks (*e.g.*, Broomhead and Lowe, 1988; Chinellato et al., 2011; Deneve et al., 2001; Deneve and Pouget, 2003; Kim et al., 2005; Latham et al., 2003; Marjanović et al., 1996; Meng and Lee, 2007, 2008; Molina-Vilaplana et al., 2004; Park and Sandberg, 1991; Pouget et al., 2003, 2002; Pouget and Sejnowski, 1994, 1997; Pouget and Snyder, 2000; Salinas and Abbott, 1995; Salinas and Sejnowski, 2001; Schilling et al., 2001; Sun and Scassellati, 2005; van Rossum and Renart, 2004; Weber et al., 2007; Weber and Wermter, 2007; Zhang et al., 2005), the PC/BC-DIM algorithm can perform sensory-sensory and motor-sensory mappings when multiple stimuli are present, as illustrated in fig. 14. However, a more sophisticated method of population decoding than is presented in sect. 2.4 would be required to select the head-centred positions corresponding to the individual targets.



(a) Step 1: mapping the visual location of the targets (R) and the eye position (E) to the head-centred locations of the targets (H)

(b) Step 2: mapping the head-centred location of one target (H) and the required retinal location of this target (R) to the required eye position (E), *i.e.*, the planned movement

(c) Step 3: mapping the head-centred locations of the targets (H) and required eye position (E) to the retinal input (R) expected after the planned movement

Figure 14: Planning gaze shifts in head-centred space with multiple visual targets. The format is the same as that used in the lower row of fig. 6, and is described in the caption of that figure. The same three steps are shown here, as are shown in fig. 6, except that the visual input encodes two targets, as indicated by the bi-modal distribution presented to the first partition in (a). In addition, the head-centred position of the target used as input to the third partition of the network in (b) is defined as -30^o rather than being calculated from the output of the first step. -30^o corresponds to the head-centred location of the visual target with the higher amplitude (the more salient one of the two visual targets). It is assumed that a more sophisticated decoding method than that presented in sect. 2.4 could determine this value from the output of the head-centred reconstruction neurons produced in the first step (a) in order to determine the input to the third partition of the network in the second step (b).

3.2.3 Stimulus Mis-Localisation and Peri-Saccadic Compression

A visual stimulus that appears briefly around the time of a saccade is not perceived at its actual position (Hamker et al., 2011; Honda, 1991; Kaiser and Lappe, 2004; Matin and Pearce, 1965; Melcher and Colby, 2008; Morrone et al., 1997). These stimulus mis-localisation effects can be divided into two types (Hamker et al., 2011; Lappe et al., 2000): peri-saccadic compression and peri-saccadic shift. Peri-saccadic shifts cause the perceived position of the briefly presented target to be shifted in the direction of the saccadic eye movement (Hamker et al., 2011; Matin and Pearce, 1965). In contrast, peri-saccadic compression results in the perceived position being moved closer to the location of the target of the saccade (Hamker et al., 2011; Lappe et al., 2000). Peri-saccadic shifts are dominant when experiments are performed in darkness, whereas compression is more significant when the experiment is perform under illuminated conditions where static landmarks are visible (Hamker et al., 2011; Lappe et al., 2000). The PC/BC-DIM model does not account for peri-saccadic shifts but does successfully simulate compression effects, although not the time-course of these effects.

To simulate peri-saccadic compression, the PC/BC-DIM network was first used to plan the upcoming saccade by following the steps described in fig. 6. It is assumed that the planned eye movement is stored in working memory for later execution, as described in sect. 3.2.2. The network then continued to perform (in a loop) sensory-sensory mappings to calculate the head-centred representation (step 1), and remappings to calculate the predicted retinal input (step 3), until after the memory-guided saccade was performed. During the period between the saccade planning and the eye movement, the probe stimulus was briefly presented as visual input to the network. Following the eye movement the network's estimate of the location of the probe stimulus was determined by decoding the head-centred representation encoded by the reconstruction neurons. Figure 15a shows these estimates of the probe location for four different probe locations.

Once the PC/BC-DIM network has been used to plan the saccade, the active prediction neurons represent the location of the saccade target. When the second stimulus is presented the activity in the network changes to reflect the location of the new stimulus. However, if the new stimulus is only present briefly, then the network does not have time to fully change state to correctly represent the new position. Instead, the position represented by the network will be intermediate between the saccade target location and the new stimulus location. This will appear as an apparent compression of visual space towards the location of the saccade target. If the probe is presented for a longer duration, then the network has more time to update the prediction neurons, and become a more accurate representation of the probe location (see fig. 15a). This is consistent with the biological data. The experimental procedure for performing a double-step saccade and for investigating peri-saccadic compression is essentially identical. However, in double-step tasks the duration of the second saccade target (which is analogous

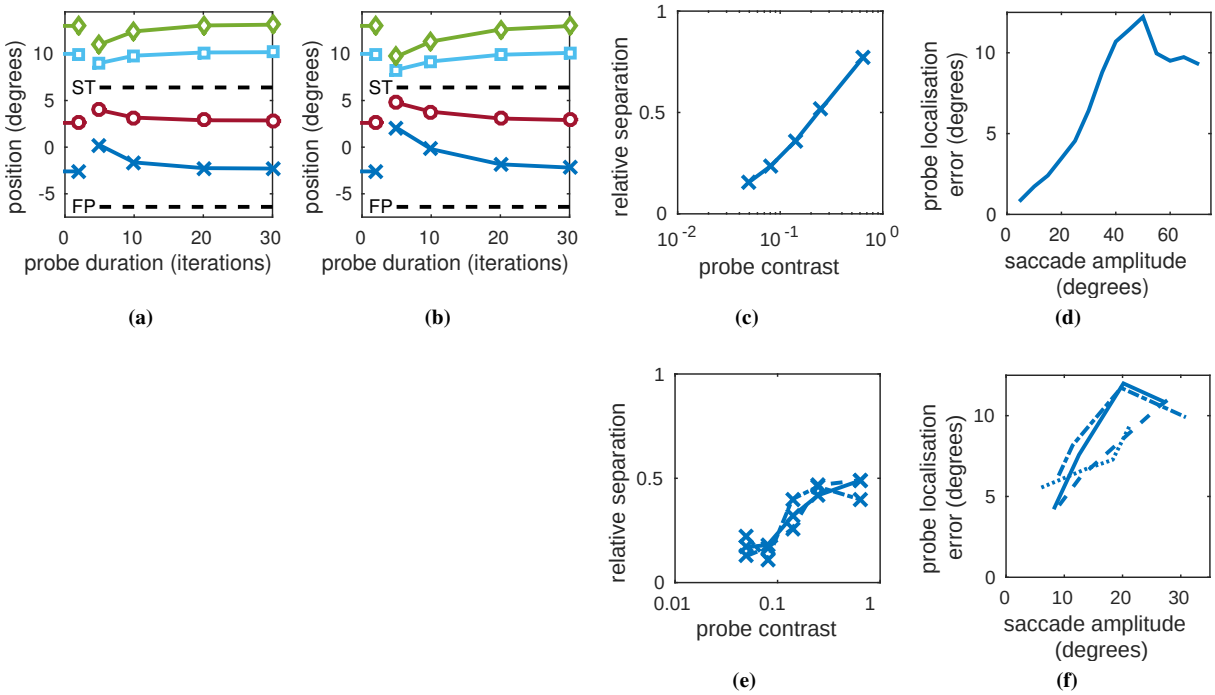


Figure 15: Peri-saccadic compression. (a) Simulation of the experiment described in (Lappe et al., 2000). The dashed lines indicate the location of the fixation point (FP), and the location of the saccade target (ST). The symbols on the left-hand edge indicate the true location at which the probe appeared. The remaining points indicate the PC/BC-DIM network's estimate of the probe locations for different probe durations. (b) As (a) but with lower contrast probes, simulated by using lower amplitude population codes to represent the visual input presented to the first partition of the PC/BC-DIM network. (c) Compression, measured using "relative separation", as a function of probe contrast. (e) Similar results for three human subjects (adapted from Fig. 4 of Michels and Lappe, 2004). (d) Compression as a function of saccade amplitude. (f) Similar results for four human subjects (adapted from Fig. 10 of Wetter and Opstal, 2008).

to the probe stimulus) is typically between 50ms and 100ms (Heide et al., 1995; Kusunoki and Goldberg, 2003; Mays and Sparks, 1980; Nakamura and Colby, 2002). For such long durations, the location of the second stimulus is accurately represented as demonstrated by the accurate second saccades made in double-step tasks. In contrast, in experiments that demonstrate compression, the probe duration is typically between 8ms and 15ms (Lappe et al., 2000; Michels and Lappe, 2004; Morrone et al., 1997; Wetter and Opstal, 2008). For such short durations, subjects perceive the probe as being shifted in position towards the location of the target for the saccade.

The explanation for peri-saccadic compression offered by the PC/BC-DIM model is distinct from that made by other models. Rather than proposing that peri-saccadic spatial compression results from shifts in retinotopic representations of visual space (Hamker et al., 2008; Pola, 2011), PC/BC-DIM proposes that compression results from a partial update of the head-centred location of the flashed target.

The speed at which the prediction neurons change to represent the location of a new visual stimulus will also be affected by stimulus contrast. A stimulus shown at a weaker contrast, simulated using a lower amplitude population code, will have a weaker influence on the prediction neurons, and hence, will be slower to change the location represented by those neurons. This is illustrated in fig. 15b where the experiment in fig. 15a is repeated using probes with half the amplitude. Michels and Lappe (2004) explicitly explored the influence of probe contrast on the magnitude of the compression effect with human subjects. PC/BC-DIM was used to simulate this data using the same location for the fixation point (-10^o), the same saccade target location ($+10^o$), and the same probe positions: -0.4^o, 5.9^o, 14.9^o, 20.4^o. The degree of compression was measured using "relative separation" which Michels and Lappe (2004) define as the standard deviation of the perceived probe locations divided by the standard deviation of the true probe positions. This measure thus ranges from 0 (when all the probes are perceived to be at the same location), to 1 (when the probes are perceived veridically). The relationship between stimulus contrast and population code amplitude is inherently arbitrary. However, for simplicity, the amplitude of the population code used for the retinal input in the model was made equal to the half the stimulus contrast used in

the experiments with human observers. In common with the results of Michels and Lappe (2004) the model shows an approximately linear relationship between probe contrast and relative suppression, with the smallest values of relative separation (the most peri-saccadic compression) occurring at the lowest contrast (fig. 15c).

In peri-saccadic compression experiments performed on human subjects, Wetter and Opstal (2008) found that the magnitude of the mis-localisation increased with the amplitude of the saccade, but that the mis-localisation error saturated at around 10^o degrees for saccades of $\geq 20^o$. Such saturation of the localisation error also occurs in the PC/BC-DIM model, as illustrated in fig. 15d. However, in the model saturation occurs for much larger saccade amplitudes than was the case with humans. In the model this is due to reaching the limit of the range of possible eye movements at about 40^o.

3.3 Neurophysiological Data

The experiments performed in the preceding sections have simulated behavioural data. The following sections will consider how well the PC/BC-DIM model accounts for neurophysiological findings related to eye movements.

3.3.1 Gain Modulation

Along the dorsal visual pathway different cortical regions represent the locations of visual stimuli in a number of distinct coordinate systems (Battaglia-Mayer et al., 2003; Blangero, 2008; Colby, 1998; Marzocchi et al., 2008; McGuire and Sabes, 2009; Pertzov et al., 2011; Weber et al., 2007), including: retinotopic (Hartmann et al., 2011), head-centred (Andersen et al., 1985; Duhamel et al., 1997; Galletti et al., 1993), body-centred (Brotchie et al., 1995; Hadjidimitrakis et al., 2013), and world-centred (Snyder et al., 1998). This is consistent with the full model illustrated in fig. 5b, in which sub-populations of the reconstruction neurons represent visual stimuli in retinotopic (stage 1, partition 1), head-centred (stage 1, partition 3 and stage 2, partition 1), body-centred (stage 2, partition 3 and stage 3, partition 1), and world-centred (stage 3, partition 3) coordinates. In the model, for convenience, the neurons in these partitions are arranged in orderly, topological, maps and are segregated from the neurons in other partitions and from the prediction neurons. However, there is no functional requirement for such segregation and orderliness, which may not therefore be observed in cortex.

Neurons that have visual responses which are modulated by postural information, such as eye or neck orientation, have also been found in many different cortical regions (*e.g.*, Andersen et al., 1990, 1985; Andersen and Mountcastle, 1983; Boussaoud et al., 1998; Bremmer, 2000; Bremmer et al., 1997; Cassanello and Ferrera, 2007; Galletti and Battaglini, 1989; Galletti et al., 1995; Guo and Li, 1997; Lehky et al., 2008; Schlag et al., 1992; Siegel et al., 2003). This is consistent with the prediction neuron responses in the PC/BC-DIM model which exhibit such gain-modulated responses. Figure 16 shows a particular example of a prediction neuron that has a retinal response preference that is modulated by eye position. A more detailed account of how PC/BC-DIM can model gain modulation has been presented previously (De Meyer and Spratling, 2011). The gain modulated responses in the PC/BC-DIM model are a consequence of performing transformations from one reference frame to another using basis-function-like neurons, and is a feature shared by many other basis-function, and non-basis-function, neural network models (*e.g.*, Deneve et al., 2001; Pouget et al., 2002; Pouget and Sejnowski, 1997; Pouget and Snyder, 2000; Salinas and Abbott, 1995, 1996; Salinas and Sejnowski, 2001; Weber et al., 2007; Xing and Andersen, 2000b; Zipser and Andersen, 1988).

3.3.2 RF Remapping and Perceptual Stability

Eye movements have been observed to alter the receptive field properties of visually responsive neurons in many brain regions (Melcher and Colby, 2008). To simulate these observations a similar procedure was used to that described in sect. 3.2.3 to simulate peri-saccadic compression. Specifically, the network shown in fig. 5c was used to perform (in a loop) sensory-sensory mappings to calculate the head-centred representation (step 1 in fig. 6), and remappings to calculate the predicted retinal input (step 3 in fig. 6). Each step was performed for 100 iterations of the PC/BC-DIM algorithm. During this time a visual stimulus was presented to the network and a memory-guided saccade was performed. The response of a single prediction neuron was recorded for a variety of different experimental conditions. Consistent with the biological experiments, but unlike some previous models (Xing and Andersen, 2000a), the recorded prediction neuron had an RF at a location that was not the same as the pre-saccade fixation point nor the target location for the saccade.

Recordings made in various cortical regions of Macaca mulatta monkeys have shown that some neurons have a pattern of response called predictive remapping (Duhamel et al., 1992; Kusunoki and Goldberg, 2003; Melcher and Colby, 2008; Nakamura and Colby, 2002; Umeno and Goldberg, 1997). In these neurons, a visual stimulus that is presented in the location that will be occupied by the RF after a saccade (the "future RF") induces a response before the saccade brings the stimulus into the RF, or induces a response that has a reduced latency following the

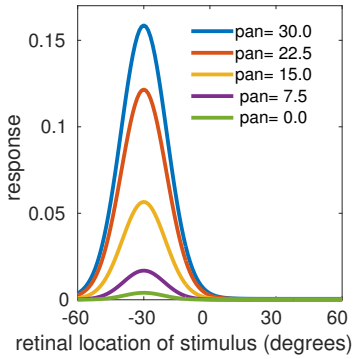


Figure 16: Gain modulated response. The response of a single prediction neuron in the network shown in fig. 5c is plotted as a function of the retinal position of the visual stimulus (R), for a number of different values of eye pan (E).

saccade (Duhamel et al., 1992; Kusunoki and Goldberg, 2003; Nakamura and Colby, 2002; Umeno and Goldberg, 1997). This is also true of prediction neurons in the PC/BC-DIM model. The second row of fig. 17a shows that the response to a stimulus in the future RF starts just prior to the saccade. This occurs in the model due to the remapping step illustrated in fig. 6c. This causes the model to predict the retinal input that is expected following the upcoming saccade. The input to the network is the head-centred representation of the stimulus (that has previously been calculated by the network) and the future eye position (which could be provided by a corollary discharge of the eye movement command, as is believed to be the cause of remapping in the cortex (Cavanaugh et al., 2016; Melcher and Colby, 2008; Wurtz, 2008)). The prediction of the upcoming visual input performed during the remapping step results in the recorded prediction neuron, which represents that combination of head-centred and eye pan signals, becoming active. Recording from the prediction neuron when the saccade is executed in the absence of any visual stimulus, as shown in the fourth row of fig. 17a, demonstrates that the remapping response is not caused by the saccade itself, as is also true of cortical cells (Umeno and Goldberg, 1997).

Another effect of eye movements on the response properties of cortical neurons is a saccade induced truncation of the response to a stimulus presented in the location occupied by the stimulus prior to the saccade (the "current RF") (Duhamel et al., 1992; Kusunoki and Goldberg, 2003; Nakamura and Colby, 2002). This is also observed in the model, as can be seen by comparing the third row with the first row of fig. 17a. The first row shows the response of a prediction neuron when a stimulus is presented to the current RF in the absence of a saccade. As with cortical neurons (Duhamel et al., 1992; Kusunoki and Goldberg, 2003), the prediction neuron has a sustained response to the stimulus even after it has disappeared. This is due to the same sub-set of prediction neurons remaining active until the network is placed in a new state by new inputs, rather like in the model of working memory shown in fig. 13. The third row of fig. 17a shows the response of the same prediction neuron when the appearance of the stimulus in its current RF is followed by a saccade. Just prior to the saccade, the response to the stimulus in the current RF is suppressed. This occurs in the network due to the remapping step (just prior to the saccade) placing the network in (and the steps following the saccade keeping the network in) a state in which a different sub-set of prediction neurons represent the visual target when it is at a new retinal position.

Predictive remapping has been shown to occur even if the stimulus presented to the future RF has disappeared before the saccade is performed (Duhamel et al., 1992; Kusunoki and Goldberg, 2003). This is particularly remarkable as it means that a neuron responds to a stimulus that is never present in its RF (Colby and Goldberg, 1999; Melcher and Colby, 2008). The PC/BC-DIM model also accounts for these results as illustrated in fig. 17b. In the lower rows of this figure it can be seen that a stimulus appearing in the future RF, but which has disappeared before the saccade, produces a weak response from the prediction neuron following the saccade. This is again due to the model acting like a working memory. The briefly presented stimulus causes a certain sub-set of prediction neurons (not including the recorded neuron) to be activated. These neurons remain active until new inputs are presented to the network. At the time of the saccade, a new value of eye position is input to the network, causing a new set of prediction neurons (that include the recorded neuron) to become active. These neurons represent the head-centred location of the previous stimulus and the new eye position.

Performing a similar experiment to that shown in fig. 17b, but where the stimulus is briefly presented in the current RF shows that the saccade induced response truncation also occurs for the sustained response to a briefly presented stimulus, see fig. 17c. This is also consistent with the behaviour of cortical neurons (Kusunoki and Goldberg, 2003). The time difference between the stimulus presentation and the saccade is negative for stimuli

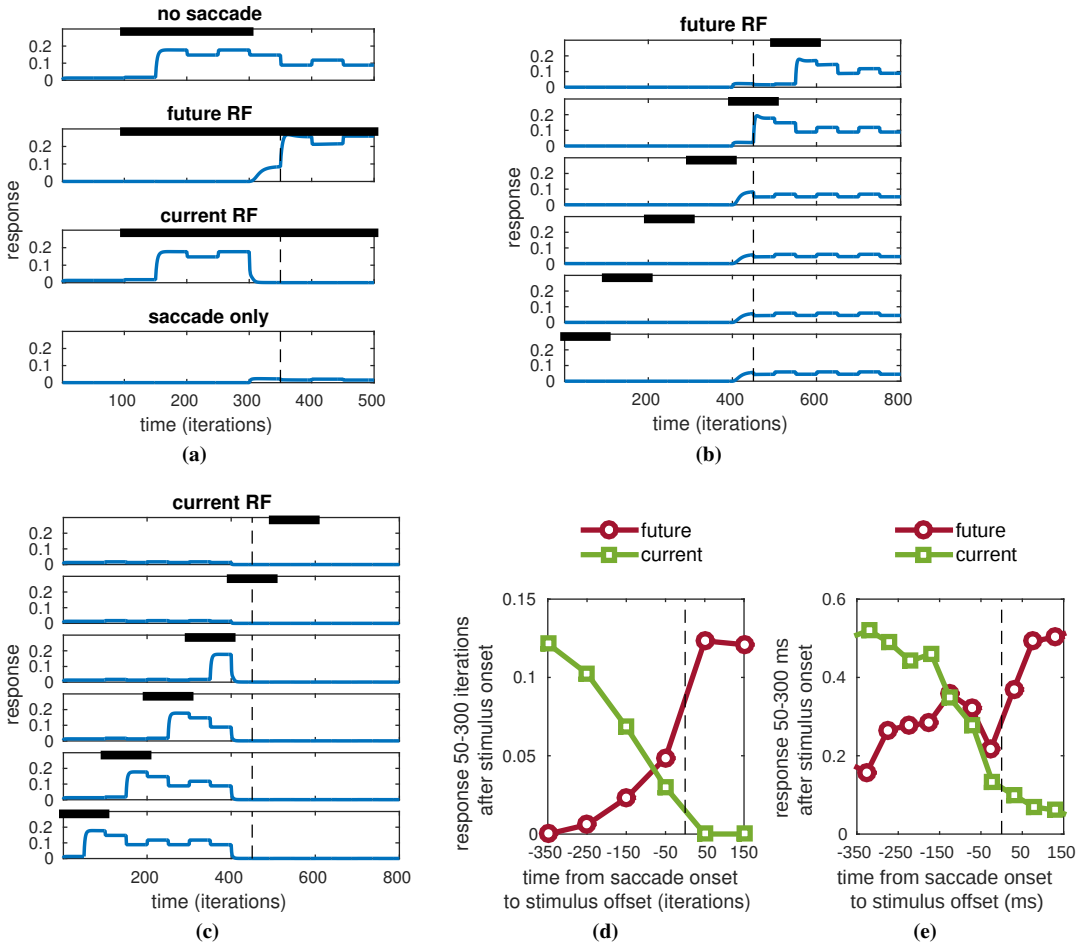


Figure 17: Simulations of RF remapping phenomena. In figures (a)-(c) the solid black line indicates the times at which the stimulus was presented. In figures (a)-(d) the vertical dashed line indicates the time at which the saccade occurred. (a) The top row shows the response of a prediction neuron to a stimulus presented within its RF in the absence of any eye movements. The second row shows the response of the same neuron when a stimulus is presented at the location occupied by the RF after a saccade. The third row shows the response of the same neuron when the stimulus is presented at the location occupied by the RF prior to the saccade. Similar results for monkey are shown in Fig. 2 of Duhamel et al. (1992). The last row shows a control condition in which the saccade is performed in the absence of a visual stimulus. Similar results for monkey are shown in Fig. 5c of Umeno and Goldberg (1997). (b) The response of the neuron to a stimulus presented at the location occupied by the RF after the saccade for different timings of the stimulus presentation relative to the saccade. Similar results for monkey are shown in Fig. 2b of Kusunoki and Goldberg (2003). (c) The response of the neuron to a stimulus presented at the location occupied by the RF prior to the saccade for different timings of the stimulus presentation relative to the saccade. Similar results for monkey are shown in Fig. 2a of Kusunoki and Goldberg (2003). (d) The data presented in (b) and (c) re-plotted to show the average response to stimuli in the future and current RF as a function of stimulus presentation time relative to the saccade. (e) Similar results for monkey (adapted from Fig. 7 of Kusunoki and Goldberg, 2003). Note that in the model saccades are instantaneous unlike in the biological data where a saccade has a small, but finite, duration (of approximately 50ms).

presented prior to the saccade and positive for stimuli presented after the saccade. As this time difference becomes less negative, so the response of the neuron to a stimulus at the pre-saccadic RF location become weaker (fig. 17c). In contrast, as the time difference becomes less negative, so the response of the neuron to a stimulus at the post-saccadic RF location become stronger (fig. 17b). This shift in responsiveness is summarised in fig. 17d, and these results are also similar to those obtained from the single-cell recordings in monkey cortex (Kusunoki and Goldberg, 2003).

There has been much debate about whether remapping is a consequence of calculations performed within a

retinotopic coordinate system (Dominey and Arbib, 1992; Droulez and Berthoz, 1991; Krommenhoek et al., 1993; Mender, 2014; Mender and Stringer, 2015; Quaia et al., 1998; Smith and Crawford, 2001; Wurtz, 2008; Xing and Andersen, 2000a), or if it is performed using neurons with responses that are partially invariant (gain-modulated cells) or fully invariant (real-position cells) to eye movement (Krommenhoek et al., 1993; Schneegans, S.and Schöner, 2012; Smith and Crawford, 2005; Ziesche and Hamker, 2014). The PC/BC-DIM model falls in the latter category. Models that employ gaze-invariant representations have an advantage in that they directly account for the perceived stability of the external world despite frequent gaze shifts.

Remapping is often presented as a mechanism that can account for the stability of visual perception in models that do not employ a gaze-invariant reference frame (Bays and Husain, 2007; Melcher and Colby, 2008; Wurtz, 2008). The idea is that the predicted input, generated through remapping, is compared to the actual input after the saccade. If the two agree then the external scene is perceived as stable (Bays and Husain, 2007). However, even with predictive remapping, neurons with retinocentric RFs seem ill positioned to solve the problem. Specifically, before and after a saccade a different sub-set of neurons will be active in such a retinotopic map. The issue is how is it possible to perceive the world as stable when completely different neurons respond to the same stable stimuli from one moment to the next. Remapping does not solve this problem. It allows the new set of neurons to respond earlier, but still results in a completely different representation of the stimulus after remapping as prior to remapping. Perceptual stability therefore seems to be better explained by a gaze-invariant representation of visual space where the same neurons respond to the same visual targets both before and after a saccade. Further support for this view comes from evidence that the stability of visual perception is impaired by parietal cortex damage (Heide et al., 1995), given that gaze-invariant representations are found in parietal regions of the cortex (Andersen et al., 1985; Brotchie et al., 1995; Duhamel et al., 1997; Galletti et al., 1993; Hadjidimitrakis et al., 2013; Snyder et al., 1998).

4 Discussion

PC/BC-DIM has previously been shown to explain a great deal of neurophysiological and psychophysical data associated with the perceptual and cognitive functions of cortex (De Meyer and Spratling, 2011, 2013; Spratling, 2008a, 2010, 2011, 2012a,b,c, 2013, 2014b, 2016b). The uniformity of cortical anatomy and physiology (Barlow, 1994; Crick and Asanuma, 1986; Douglas et al., 1989; Ebdon, 1992, 1996; Harris and Shepherd, 2015; Mountcastle, 1998; Mumford, 1992, 1994; Phillips and Singer, 1997) suggests that the same mechanisms that underly perception and cognition should also be involved in action. To determine whether or not PC/BC-DIM can account for neurophysiological and psychophysical data associated with action, it has been applied in this article to simulating gaze shifts. The prevalence of movements to control gaze in primates, and the abundance of behavioural and electro-physiological data concerning eye and gaze movements, makes gaze shifts a sensible first target for testing a PC/BC-DIM based model of action.

The proposed model shares many features in common with previous models. Specifically, the sensory and motor variables are encoded using Gaussian population codes, the prediction neurons act as basis functions encoding the relationships between the sensory and motor variables, and the actions are planned using gaze-invariant representations of visual space. The correspondence between PC/BC-DIM and basis function networks has been noted previously (De Meyer and Spratling, 2011; Muhammad and Spratling, 2015; Spratling, 2009, 2012c, 2016a,b). The current work, and other recent work applying the PC/BC-DIM model to robot control (Muhammad and Spratling, 2017a,b), goes beyond this previously observed correspondence by showing that PC/BC-DIM can be used to plan coordinated movements of multiple DOFs to bring about an action. Furthermore, the current results show that the relationships between eye, head and body movements during coordinated gaze shifts are consistent with those observed in primates (sect. 3.1). When planning coordinated movements, the model calculates the required position of the eye first, then the neck, and finally the torso. This order is consistent the sequential onset of these different movements in primates (McCluskey and Cullen, 2007; Tweed et al., 1995). It is this sequential order of movement planning, together with weak inputs representing the current posture, that results in the model favouring smaller body and neck movements over large ones and produces movements of eye, head and body that are consistent with the stereotypical movements observed in primates (Freedman and Sparks, 1997; McCluskey and Cullen, 2007). This method of resolving kinematic redundancies in the PC/BC-DIM model is quite distinct from those methods proposed in previous models (Harris and Wolpert, 1998, 2006; Kardamakis and Moschovakis, 2009; Saeb et al., 2011; Todorov, 2004).

The use of population coding to represent sensory and motor variables is important to allow the model to accurately represent continuous values using a finite number of neurons. This allows the PC/BC-DIM model to simulate saccades with an accuracy consistent with primates (sect. 3.2.1). In addition, the use of population coding naturally results in saturation of eye movements before they reach the physical limits. This saturation

also contributes to the accurate simulation of stereotypical combinations of movements observed in primates (Freedman and Sparks, 1997; McCluskey and Cullen, 2007, sect. 3.1), and the undershoot of large magnitude saccades (sect. 3.2.1) as is also observed in primates (Albano and Wurtz, 1982; Leigh and Kennard, 2004).

Using the PC/BC-DIM network to act like a basis-function network is necessary to allow the proposed model to perform mappings between different sensory and motor variables, and coordinate transformations more generally. However, to make such a solution tractable when there are many sensory and motor variables it is necessary to decompose transformations into multiple steps, each of which is implemented by a separate basis-function network (Deneve and Pouget, 2003; Pouget and Sejnowski, 1997). This sub-division of complex transformations requires internal variables to facilitate the coordination of the separate networks. These internal variables naturally represent visual targets in reference frames that are invariant to movement. The presence of retina-centred, head-centred, body-centred, and world-centred representations of visual space in the proposed model is consistent with the existence of multiple coordinate systems along the dorsal pathway of the cortical visual system (Battaglia-Mayer et al., 2003; Blangero, 2008; Marzocchi et al., 2008; McGuire and Sabes, 2009; Pertzov et al., 2011), including retinotopic (Hartmann et al., 2011), head-centred (Andersen et al., 1985; Duhamel et al., 1997; Galletti et al., 1993), body-centred (Brotchie et al., 1995; Hadjidimitrakis et al., 2013), and world-centred (Snyder et al., 1998) representations of visual stimuli. The model proposes that these multiple coordinate systems are a natural consequence of representing mappings between many sensory and motor variable in a tractable way using basis-function style neural networks.

The presence in the PC/BC-DIM model of reference frames that are invariant to posture provides a simple explanation for the perception of a stationary world (Krommenhoek et al., 1993; Schneegans, S.and Schöner, 2012; Smith and Crawford, 2005; Ziesche and Hamker, 2014, sect. 3.3.2). It also allows the model to explain the ability to perform accurate saccades despite previous movements, as occurs in double-step saccade experiments (Mays and Sparks, 1980; Robinson, 1975, sect. 3.2.2). The proposed PC/BC-DIM model can also act as a working memory to enable the locations of targets to be remembered, as would be required for memory-guided gaze shifts like double-step saccades. This ability of the model to act as a working memory, combined with the ability of the model to use a gaze-invariant representation of a visual target to predict the retinal input that is expected following an upcoming saccade, also enables the model to successfully simulate RF remapping (Duhamel et al., 1992; Kusunoki and Goldberg, 2003; Umeno and Goldberg, 1997, sect. 3.3.2). In double-step saccade experiments, the second stimulus is presented for a sufficiently long time that the location of this target can be accurately represented by the PC/BC-DIM network. However, if the second target is more briefly presented, then the PC/BC-DIM network does not have sufficient time to update the representation fully, resulting in the location of the second target encoded by the network being intermediate between its true location and the location of the preceding visual target. This behaviour allows the proposed network to account peri-saccadic compression: a shift in the perceived location of a stimulus towards the saccade target (Lappe et al., 2000; Michels and Lappe, 2004; Wetter and Opstal, 2008, sect. 3.2.3). This explanation of peri-saccadic compression is quite distinct from those proposed in previous models (Hamker et al., 2008; Pola, 2011).

A final consequence of using prediction neurons to act as basis functions encoding the relationships between the sensory and motor variables, is that these neurons have gain-modulated responses (sect. 3.3.1) consistent with the existence of cortical neurons that have visual responses modulated by posture (Andersen et al., 1990, 1985; Andersen and Mountcastle, 1983; Boussaoud et al., 1998; Bremmer, 2000; Bremmer et al., 1997; Cassanello and Ferrera, 2007; Galletti and Battaglini, 1989; Galletti et al., 1995; Guo and Li, 1997; Lehky et al., 2008; Schlag et al., 1992; Siegel et al., 2003).

Despite these successes the current model does fail to account for some phenomena related to eye movements. Specifically, suppression of awareness to visual blur during saccades (saccadic suppression); shifts in the perceived location of a stimulus in the direction of a saccadic eye movement (peri-saccadic shifts); and, the failure of the visual system to detect changes in stimulus location made during a saccade (suppression of displacement). The current model also does not consider learning or adaptation, and hence, does not address changes in the amplitudes of saccades that are induced by persistent mismatches between the predicted and actual post-saccadic location of a target (saccadic adaptation). The proposed model also does not address the dynamics of gaze control, or visual pursuit or tracking movements. Furthermore, the current model only offers an incomplete account of those phenomena which it does successfully simulate. For a more complete explanation there is a need for additional mechanisms, not simulated here, that can coordinate the overall process: deciding what type of mapping to perform, when to execute actions, when to store information in working memory and when to recall it. The current model also does not deal with how to select where to shift gaze when there are multiple potential targets (Henderson, 2017). The current model is also rather crude in assuming that coordinate transformations are performed in sequence and that each is performed for a set period of time. Having been applied to perception, cognition, and now action planning it would be interesting to see if predictive coding could also be extended to implement the type of executive functions required to build a more complete model.

Disclosure of interest

The author reports no conflicts of interest.

References

Adams, R., Shipp, S., and Friston, K. (2013). Predictions not commands: active inference in the motor system. *Brain Structure and Function*, 218(3):611–43.

Albano, J. E. and Wurtz, R. H. (1982). Deficits in eye position following ablation of monkey superior colliculus, pretectum, and posterior-medial thalamus. *Journal of Neurophysiology*, 48(2):318–37.

Andersen, R. A., Bracewell, R. M., Barash, S., Gnadt, J. W., and Fogassi, L. (1990). Eye position effects on visual, memory, and saccade-related activity in areas LIP and 7a of macaque. *The Journal of Neuroscience*, 10:1176–96.

Andersen, R. A., Essick, G. K., and Siegel, R. M. (1985). Encoding of spatial location by posterior parietal neurons. *Science*, 230(4724):456–8.

Andersen, R. A. and Mountcastle, V. B. (1983). The influence of the angle of gaze upon the excitability of the light-sensitive neurons of the posterior parietal cortex. *The Journal of Neuroscience*, 3(3):532–48.

Barlow, H. B. (1994). What is the computational goal of the neocortex? In Koch, C. and Davis, J. L., editors, *Large-Scale Neuronal Theories of the Brain*, chapter 1, pages 1–22. MIT Press, Cambridge, MA.

Battaglia-Mayer, A., Caminiti, R., Lacquaniti, F., and Zago, M. (2003). Multiple levels of representation of reaching in the parieto-frontal network. *Cerebral Cortex*, 13(10):1009–22.

Bays, P. M. and Husain, M. (2007). Spatial remapping of the visual world across saccades. *NeuroReport*, 18:1207–13.

Becker, W. and Jürgens, R. (1979). An analysis of the saccadic system by means of double step stimuli. *Vision Research*, 19:967–83.

Blangero, A. (2008). *The Sensorimotor Functions Of The Posterior Parietal Cortex: Evidence from patients with Optic Ataxia*. PhD thesis, L'Universite De Lyon, France.

Boussaoud, D., Jouffrais, C., and Bremmer, F. (1998). Eye position effects on the neuronal activity of dorsal premotor cortex in the macaque monkey. *Journal of Neurophysiology*, 80:1132–50.

Bremmer, F. (2000). Eye position effects in macaque area V4. *NeuroReport*, 11:1277–83.

Bremmer, F., Distler, C., and Hoffmann, K. P. (1997). Eye position effects in monkey cortex. II. pursuit- and fixation-related activity in posterior parietal areas LIP and 7a. *Journal of Neurophysiology*, 77(2):962–77.

Broomhead, D. S. and Lowe, D. (1988). Multivariable functional interpolation and adaptive networks. *Complex Systems*, 2:321–55.

Brotchie, P. R., Andersen, R. A., Snyder, L. H., and Goodman, S. J. (1995). Head position signals used by parietal neurons to encode locations of visual stimuli. *Nature*, 375(6528):232–5.

Bubic, A., von Cramon, D. Y., and Schubotz, R. I. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience*, 4(25):1–15.

Cassanello, C. R. and Ferrera, V. P. (2007). Computing vector differences using a gain field-like mechanism in monkey frontal eye field. *Journal of Physiology*, 582(2):647–64.

Cavanaugh, J., Berman, R. A., Joiner, W. M., and Wurtz, R. H. (2016). Saccadic corollary discharge underlies stable visual perception. *The Journal of Neuroscience*, 36(1):31–42.

Chinellato, E., Antonelli, M., Grzyb, B., and del Pobil, A. P. (2011). Implicit sensorimotor mapping of the peripersonal space by gazing and reaching. *IEEE Transactions on Autonomous Mental Development*, 3(1):43–52.

Clark, A. (2013). Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(03):181–204.

Cohen, M. E. and Ross, L. E. (1978). Latency and accuracy characteristics of saccades and corrective saccades in children and adults. *Journal of Experimental Child Psychology*, 26(3):517–27.

Colby, C. L. (1998). Action-oriented spatial reference frames in cortex. *Neuron*, 20:15–24.

Colby, C. L. and Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annual Review of Neuroscience*, 22(1):319–49.

Crick, F. and Asanuma, C. (1986). Certain aspects of the anatomy and physiology of the cerebral cortex. In Rumelhart, D. E., McClelland, J. L., and The PDP Research Group, editors, *Parallel Distributed Processing: Explorations in the Microstructures of Cognition. Volume 2: Psychological and Biological Models*, pages 333–71. MIT Press, Cambridge, MA.

De Meyer, K. and Spratling, M. W. (2011). Multiplicative gain modulation arises through unsupervised learning in a predictive coding model of cortical function. *Neural Computation*, 23(6):1536–67.

De Meyer, K. and Spratling, M. W. (2013). A model of partial reference frame transforms through pooling of gain-modulated responses. *Cerebral Cortex*, 23(5):1230–9.

Deneve, S., Latham, P. E., and Pouget, A. (2001). Efficient computation and cue integration with noisy population codes. *Nature Neuroscience*, 4(8):826–31.

Deneve, S. and Pouget, A. (2003). Basis functions for object-centered representations. *Neuron*, 37:347–59.

Dominey, P. F. and Arbib, M. A. (1992). A cortico-subcortical model for generation of spatially accurate sequential saccades. *Cerebral Cortex*, 2(2):153–75.

Douglas, R. J., Martin, K. A. C., and Whitteridge, D. (1989). A canonical microcircuit for neocortex. *Neural Computation*, 1(4):480–8.

Droulez, J. and Berthoz, A. (1991). A neural network model of sensoritopic maps with predictive short-term memory properties. *Proceedings of the National Academy of Sciences USA*, 88:9653–7.

Duhamel, J., Colby, C., and Goldberg, M. (1992). The updating of the representation of visual space in parietal cortex by intended eye movements. *Science*, 255:90–2.

Duhamel, J.-R., Bremmer, F., BenHamed, S., and Graf, W. (1997). Spatial invariance of visual receptive fields in parietal cortex neurons. *Nature*, 389:845–8.

Ebdon, M. (1992). The uniformity of cerebral neocortex and its implications for cognitive science. Technical Report CSRP-228, School of Cognitive and Computing Sciences, University of Sussex.

Ebdon, M. (1996). *Towards a General Theory of Cerebral Neocortex*. PhD thesis, University of Sussex, UK.

Findlay, J. and Walker, R. (2012). Human saccadic eye movements. *Scholarpedia*, 7(7):5095. revision 122018.

Flash, T. and Sejnowski, T. J. (2001). Computational approaches to motor control. *Current Opinion in Neurobiology*, 11(6):655–62.

Földiák, P. (1991). Learning invariance from transformation sequences. *Neural Computation*, 3:194–200.

Freedman, E. G. and Sparks, D. L. (1997). Eye-head coordination during head-unrestrained gaze shifts in rhesus monkeys. *Journal of Neurophysiology*, 77:2328–48.

Friston, K., Adams, R. A., Perrinet, L., and Breakspear, M. (2012). Perceptions as hypotheses: Saccades as experiments. *Frontiers in Psychology*, 3:151.

Friston, K. and Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 364:1211–21.

Friston, K., Mattout, J., and Kilner, J. (2011). Action understanding and active inference. *Biological Cybernetics*, 104(1-2):137–60.

Friston, K. J., Daunizeau, J., Kilner, J., and Kiebel, S. J. (2010). Action and behavior: a free-energy formulation. *Biological Cybernetics*, 102(3):227–60.

Galletti, C. and Battaglini, P. P. (1989). Gaze-dependent visual neurons in area V3A of monkey prestriate cortex. *The Journal of Neuroscience*, 9(4):1112–25.

Galletti, C., Battaglini, P. P., and Fattori, P. (1993). Parietal neurons encoding spatial locations in craniotopic coordinates. *Experimental Brain Research*, 96:221–9.

Galletti, C., Battaglini, P. P., and Fattori, P. (1995). Eye position influence on the parieto-occipital area PO (V6) of the macaque monkey. *European Journal of Neuroscience*, 7(12):2486–501.

Georgopoulos, A. P., Schwartz, A. B., and Kettner, R. E. (1986). Neuronal population coding of movement direction. *Science*, 233:1416–9.

Goldberg, M. E. and Bruce, C. J. (1990). Primate frontal eye fields. III. maintenance of a spatially accurate saccade signal. *Journal of Neurophysiology*, 64:489–508.

Guitton, D. and Volle, M. (1987). Gaze control in humans: eye-head coordination during orienting movements to targets within and beyond the oculomotor range. *Journal of Neurophysiology*, 58(3):427–59.

Guo, K. and Li, C.-Y. (1997). Eye position-dependent activation of neurones in striate cortex of macaque. *NeuroReport*, 8:1405–9.

Hadjidimitrakis, K., Bertozzi, F., Breveglieri, R., Fattori, P., and Galletti, C. (2013). Body-centered, mixed, but not hand-centered coding of visual targets in the medial posterior parietal cortex during reaches in 3D space. *Cerebral Cortex*, 24(12):3209–20.

Hallett, P. E. and Lightstone, A. (1976). Saccadic eye movements to flashed targets. *Vision Research*, 16(1):107–14.

Hamker, F. H., Zirnsak, M., Calow, D., and Lappe, M. (2008). The peri-saccadic perception of objects and space. *PLoS Computational Biology*, 4:e31.

Hamker, F. H., Zirnsak, M., Ziesche, A., and Lappe, M. (2011). Computational models of spatial updating in peri-saccadic perception. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 366(1564):554–71.

Harris, C. M. and Wolpert, D. M. (1998). Signal-dependent noise determines motor planning. *Nature*, 394:780–4.

Harris, C. M. and Wolpert, D. M. (2006). The main sequence of saccades optimizes speed-accuracy trade-off.

Biological Cybernetics, 95:21–9.

Harris, K. D. and Shepherd, G. M. (2015). The neocortical circuit: Themes and variations. *Nature Neuroscience*, 18:170–81.

Hartmann, T. S., Bremmer, F., Albright, T. D., and Krekelberg, B. (2011). Receptive field positions in area MT during slow eye movements. *The Journal of Neuroscience*, 31(29):10437–44.

Heide, W., Blankenburg, M., Zimmermann, E., and Kömpf, D. (1995). Cortical control of double-step saccades: Implications for spatial orientation. *Annals of Neurology*, 38(5):739–48.

Henderson, J. M. (2017). Gaze control as prediction. *Trends in Cognitive Sciences*, 21(1):15–23.

Henson, D. B. (1978). Corrective saccades: Effects of altering visual feedback. *Vision Research*, 18(1):63–7.

Hinton, G. E., Osindero, S., and Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, 18:1527–54.

Honda, H. (1991). The time courses of visual mislocalization and of extra-retinal eye position signals at the time of vertical saccades. *Vision Research*, 31:1915–21.

Houk, J. C., Buckingham, J. T., and Barto, A. G. (1996). Models of the cerebellum and motor learning. *Behavioral and Brain Sciences*, 19(3):368–83.

Huang, Y. and Rao, R. P. N. (2011). Predictive coding. *WIREs Cognitive Science*, 2:580–93.

Kaiser, M. and Lappe, M. (2004). Perisaccadic mislocalization orthogonal to saccade direction. *Neuron*, 41(2):293–300.

Kapoula, Z. and Robinson, D. A. (1986). Saccadic undershoot is not inevitable: Saccades can be accurate. *Vision Research*, 26(5):735–43.

Kardamakis, A. A. and Moschovakis, A. K. (2009). Optimal control of gaze shifts. *The Journal of Neuroscience*, 29(24):7723–30.

Kawato, M. (1995). Cerebellum and motor control. In Arbib, M. A., editor, *The Handbook of Brain Theory and Neural Networks*, pages 172–7. MIT Press, Cambridge, MA.

Kim, D., Huh, S.-H., Seo, S.-J., and Park, G.-T. (2005). Self-organizing radial basis function network modeling for robot manipulator. In Ali, M. and Esposito, F., editors, *Innovations in Applied Artificial Intelligence*, volume 3533 of *Lecture Notes in Computer Science*, pages 579–87. Springer Berlin Heidelberg.

Komoda, M. K., Festinger, L., Philips, L. J., Duckman, R. H., and Young, R. A. (1973). Some observations concerning saccadic eye movements. *Vision Research*, 13:1009–20.

Krommenhoek, K., Opstal, A. V., Gielen, C., and Gisbergen, J. V. (1993). Remapping of neural activity in the motor colliculus: a neural network study. *Vision Research*, 33:1287–98.

Kusunoki, M. and Goldberg, M. E. (2003). The time course of perisaccadic receptive field shifts in the lateral intraparietal area of the monkey. *Journal of Neurophysiology*, 89:1519–27.

Lappe, M., Awater, H., and Krekelberg, B. (2000). Postsaccadic visual references generate presaccadic compression of space. *Nature*, 403:892–5.

Larochelle, H., Bengio, Y., Louradour, J., and Lamblin, P. (2009). Exploring strategies for training deep neural networks. *Journal of Machine Learning Research*, 1:1–40.

Latham, P. E., Deneve, S., and Pouget, A. (2003). Optimal computation with attractor networks. *Journal of Physiology – Paris*, 97(4–6):683–94.

Law, J., Shaw, P., and Lee, M. (2013). A biologically constrained architecture for developmental learning of eyehead gaze control on a humanoid robot. *Autonomous Robots*, 35(1):77–92.

Lehky, S. R., Peng, X., McAdams, C. J., and Sereno, A. B. (2008). Spatial modulation of primate inferotemporal responses by eye position. *PLoS ONE*, 3:e3492.

Leigh, R. J. and Kennard, C. (2004). Using saccades as a research tool in the clinical neurosciences. *Brain*, 127(3):460–77.

Marjanović, M., Scassellati, B., and Williamson, M. (1996). Self-taught visually guided pointing for a humanoid robot. In Maes, P., Mataric, M., Meyer, J.-A., Pollack, J., and Wilson, S. W., editors, *From Animals to Animats: Proceedings of the International Conference on Simulation of Adaptive Behaviour*, pages 35–44, Cambridge, MA. MIT Press.

Marzocchi, N., Breveglieri, R., Galletti, C., and Fattori, P. (2008). Reaching activity in parietal area V6A of macaque: eye influence on arm activity or retinocentric coding of reaching movements? *European Journal of Neuroscience*, 27(3):775–89.

Matin, L. and Pearce, D. G. (1965). Visual perception of direction for stimuli flashed during voluntary saccadic eye movmements. *Science*, 148:1485–7.

Mays, L. E. and Sparks, D. L. (1980). Dissociation of visual and saccade-related responses in superior colliculus neurons. *Journal of Neurophysiology*, 43:207–32.

McCluskey, M. K. and Cullen, K. E. (2007). Eye, head, and body coordination during large gaze shifts in rhesus monkeys: Movement kinematics and the influence of posture. *Journal of Neurophysiology*, 97:2976–91.

McGuire, L. M. M. and Sabes, P. N. (2009). Sensory transformations and the use of multiple reference frames for reach planning. *Nature Neuroscience*, 12(8):1056–61.

Melcher, D. and Colby, C. L. (2008). Trans-saccadic perception. *Trends in Cognitive Sciences*, 12:466–73.

Mender, B. M. W. (2014). *Models of Primate Supraretinal Visual Representations*. PhD thesis, University of Oxford.

Mender, B. M. W. and Stringer, S. M. (2015). A self-organizing model of perisaccadic visual receptive field dynamics in primate visual and oculomotor system. *Frontiers in Computational Neuroscience*, 9(17).

Meng, Q. and Lee, M. H. (2007). Automated cross-modal mapping in robotic eye/hand systems using plastic radial basis function networks. *Connection Science*, 19(1):25–52.

Meng, Q. and Lee, M. H. (2008). Error-driven active learning in growing radial basis function networks for early robot learning. *Neurocomputing*, 71(7–9):1449–61.

Michels, L. and Lappe, M. (2004). Contrast dependency of saccadic compression and suppression. *Vision Research*, 44(20):2327–36.

Mohsenzadeh, Y., Dash, S., and Crawford, J. D. (2016). A state space model for spatial updating of remembered visual targets during eye movements. *Frontiers in Systems Neuroscience*, 10(39).

Molina-Vilaplana, J., Pedreño-Molina, J. L., and López-Coronado, J. (2004). Hyper RBF model for accurate reaching in redundant robotic systems. *Neurocomputing*, 61:495–501.

Morrone, M. C., Ross, J., and Burr, D. C. (1997). Apparent position of visual targets during real and simulated saccadic eye movements. *The Journal of Neuroscience*, 17:7941–53.

Mountcastle, V. B. (1998). *Perceptual Neuroscience: The Cerebral Cortex*. Harvard University Press, Cambridge, MA.

Muhammad, W. and Spratling, M. W. (2015). A neural model of binocular saccade planning and vergence control. *Adaptive Behavior*, 23(5):265–82.

Muhammad, W. and Spratling, M. W. (2017a). A neural model for eye-head-arm coordination. *Advanced Robotics*, 31(12):650–63.

Muhammad, W. and Spratling, M. W. (2017b). A neural model of coordinated head and eye movement control. *Journal of Intelligent & Robotic Systems*, 85(1):107–26.

Mumford, D. (1992). On the computational architecture of the neocortex II: the role of cortico-cortical loops. *Biological Cybernetics*, 66:241–51.

Mumford, D. (1994). Neuronal architectures for pattern-theoretic problems. In Koch, C. and Davis, J. L., editors, *Large-Scale Neuronal Theories of the Brain*, pages 125–52. MIT Press, Cambridge, MA.

Nakamura, K. and Colby, C. L. (2002). Updating of the visual representation in monkey striate and extrastriate cortex during saccades. *Proceedings of the National Academy of Sciences USA*, 99:4026–31.

Niemeier, M., Crawford, J. D., and Tweed, D. B. (2003). Optimal transsaccadic integration explains distorted spatial perception. *Nature*, 422(6927):76–80.

O'Reilly, R. C. and McClelland, J. L. (1992). The self-organization of spatially invariant representations. Technical Report PDP.CNS.92.5, Department of Psychology, Carnegie Mellon University.

Park, J. and Sandberg, I. W. (1991). Universal approximation using radial-basis-function networks. *Neural Computation*, 3:246–57.

Perrinet, L. U., Adams, R. A., and Friston, K. J. (2014). Active inference, eye movements and oculomotor delays. *Biological Cybernetics*, 108(6):777–801.

Pertzov, Y., Avidan, G., and Zohary, E. (2011). Multiple reference frames for saccadic planning in the human parietal cortex. *The Journal of Neuroscience*, 31(3):1059–68.

Phillips, W. A. and Singer, W. (1997). In search of common foundations for cortical computation. *Behavioral and Brain Sciences*, 20(4):657–722.

Pola, J. (2011). An explanation of perisaccadic compression of visual space. *Vision Research*, 51(4):424–34.

Pouget, A., Dayan, P., and Zemel, R. S. (2003). Inference and computation with population codes. *Annual Review of Neuroscience*, 26:381–410.

Pouget, A., Deneve, S., and Duhamel, J. R. (2002). A computational perspective on the neural basis of multisensory spatial representations. *Nature Reviews Neuroscience*, 3:741–7.

Pouget, A. and Sejnowski, T. J. (1994). A neural model of the cortical representation of egocentric distance. *Cerebral Cortex*, 4(3):314–29.

Pouget, A. and Sejnowski, T. J. (1997). Spatial transformations in the parietal cortex using basis functions. *Journal of Cognitive Neuroscience*, 9(2):222–37.

Pouget, A. and Snyder, L. (2000). Computational approaches to sensorimotor transformations. *Nature Neuroscience*, 3(supplement):1192–8.

Quaia, C., Optican, L. M., and Goldberg, M. E. (1998). The maintenance of spatial accuracy by the perisaccadic remapping of visual receptive fields. *Neural Networks*, 11:1229–40.

Rao, R. P. N. and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1):79–87.

Robinson, D. (1975). Oculomotor control signals. In Lennerstrand, G. and Bach-Y-Rita, P., editors, *Basic mechanisms of ocular motility and their clinical implications*, pages 337–78. Pergamon Press, Oxford, UK.

Saeb, S., Weber, C., and Triesch, J. (2011). Learning the optimal control of coordinated eye and head movements. *PLoS Computational Biology*, 7(11):e1002253.

Salinas, E. and Abbott, L. F. (1995). Transfer of coded information from sensory to motor networks. *The Journal of Neuroscience*, 15:6461–74.

Salinas, E. and Abbott, L. F. (1996). A model of multiplicative neural responses in parietal cortex. *Proceedings of the National Academy of Sciences USA*, 93:11956–61.

Salinas, E. and Sejnowski, T. J. (2001). Gain modulation in the central nervous system: where behavior, neurophysiology and computation meet. *The Neuroscientist*, 7(5):430–40.

Schilling, R. J., Carroll, J. J., and Al-Ajlouni, A. F. (2001). Approximation of nonlinear systems with radial basis function neural networks. *IEEE Transactions on Neural Networks*, 12(1):1–15.

Schlag, J., Schlag-Rey, M., and Pigarev, I. (1992). Supplementary eye field: influence of eye position on neural signals of fixation. *Experimental Brain Research*, 90:302–6.

Schneegans, S.and Schöner, G. (2012). A neural mechanism for coordinate transformation predicts pre-saccadic remapping. *Biological Cybernetics*, 106(2):89–109.

Siegel, R. M., Raffi, M., Phinney, R. E., Turner, J. A., and Jando, G. (2003). Functional architecture of eye position gain fields in visual association cortex of behaving monkey. *Journal of Neurophysiology*, 90:1279–94.

Smith, M. and Crawford, J. (2001). Self-organizing task modules and explicit coordinate systems in a neural network model for 3-D saccades. *Journal of Computational Neuroscience*, 10:127–50.

Smith, M. and Crawford, J. (2005). Distributed population mechanism for the 3-D oculomotor reference frame transformation. *Journal of Neurophysiology*, 93:1742–61.

Snyder, L. H., Grieve, K. L., Brotchie, P., and Andersen, R. A. (1998). Separate body- and world-referenced representations of visual space in parietal cortex. *Nature*, 394:887–91.

Spratling, M. W. (2005). Learning viewpoint invariant perceptual representations from cluttered images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5):753–61.

Spratling, M. W. (2008a). Predictive coding as a model of biased competition in visual selective attention. *Vision Research*, 48(12):1391–408.

Spratling, M. W. (2008b). Reconciling predictive coding and biased competition models of cortical function. *Frontiers in Computational Neuroscience*, 2(4):1–8.

Spratling, M. W. (2009). Learning posture invariant spatial representations through temporal correlations. *IEEE Transactions on Autonomous Mental Development*, 1(4):253–63.

Spratling, M. W. (2010). Predictive coding as a model of response properties in cortical area V1. *The Journal of Neuroscience*, 30(9):3531–43.

Spratling, M. W. (2011). A single functional model accounts for the distinct properties of suppression in cortical area V1. *Vision Research*, 51(6):563–76.

Spratling, M. W. (2012a). Predictive coding accounts for V1 response properties recorded using reverse correlation. *Biological Cybernetics*, 106(1):37–49.

Spratling, M. W. (2012b). Predictive coding as a model of the V1 saliency map hypothesis. *Neural Networks*, 26:7–28.

Spratling, M. W. (2012c). Unsupervised learning of generative and discriminative weights encoding elementary image components in a predictive coding model of cortical function. *Neural Computation*, 24(1):60–103.

Spratling, M. W. (2013). Image segmentation using a sparse coding model of cortical area V1. *IEEE Transactions on Image Processing*, 22(4):1631–43.

Spratling, M. W. (2014a). Classification using sparse representations: a biologically plausible approach. *Biological Cybernetics*, 108(1):61–73.

Spratling, M. W. (2014b). A single functional model of drivers and modulators in cortex. *Journal of Computational Neuroscience*, 36(1):97–118.

Spratling, M. W. (2016a). A neural implementation of Bayesian inference based on predictive coding. *Connection Science*, 28(4):346–83.

Spratling, M. W. (2016b). Predictive coding as a model of cognition. *Cognitive Processing*, 17(3):279–305.

Spratling, M. W. (2017). A review of predictive coding algorithms. *Brain and Cognition*, 112:92–7.

Spratling, M. W., De Meyer, K., and Kompass, R. (2009). Unsupervised learning of overlapping image components using divisive input modulation. *Computational Intelligence and Neuroscience*, 2009(381457):1–19.

Sun, G. and Scassellati, B. (2005). A fast and efficient model for learning to reach. *Int. J. Human. Robot.*, 2(4):391–413.

Templeman, J. N. and Loew, M. H. (1989). Staged assimilation: a system for detecting invariant features in temporally coherent visual stimuli. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN89)*, volume 1, pages 731–8, New York, NY. IEEE Press.

Todorov, E. (2004). Optimality principles in sensorimotor control. *Nature Neuroscience*, 7:907–15.

Tomlinson, R. D. and Bahra, P. S. (1986). Combined eye-head gaze shifts in the primate. I. metrics. *Journal of Neurophysiology*, 56(6):1542–57.

Tweed, D. (1997). Three-dimensional model of the human eye-head saccadic system. *Journal of Neurophysiology*, 77:654–66.

Tweed, D., Glenn, B., and Vilis, T. (1995). Eye-head coordination during large gaze shifts. *Journal of Neurophysiology*, 73(2):766–79.

Umeno, M. M. and Goldberg, M. E. (1997). Spatial processing in the monkey frontal eye field. I. predictive visual responses. *Journal of Neurophysiology*, 78:1373–83.

van Hemmen, J. L. and Schwartz, A. B. (2008). Population vector code: a geometric universal as actuator. *Biological Cybernetics*, 98(6):509–18.

van Rossum, M. C. W. and Renart, A. (2004). Computation with populations codes in layered networks of integrate-and-fire neurons. *Neurocomputing*, 58–60:265–70.

Wallis, G. (1996). Using spatio-temporal correlations to learn invariant object recognition. *Neural Networks*, 9(9):1513–9.

Weber, C., Elshaw, M., Triesch, J., and Wermter, S. (2007). Neural control of actions involving different coordinate systems. In Hackel, M., editor, *Humanoid Robots: Human-like Machines*. I-Tech Education and Publishing, Vienna, Austria.

Weber, C. and Wermter, S. (2007). A self-organizing map of sigma-pi units. *Neurocomputing*, 70(13-15):2552–60.

Westheimer, G. (1954). Eye movement responses to a horizontally moving visual stimulus. *Arch. Ophthol.*, 52:932–41.

Wetter, S. M. C. I. V. and Opstal, A. J. V. (2008). Experimental test of visuomotor updating models that explain perisaccadic mislocalization. *Journal of Vision*, 8(14):8.

Wolpert, D. M. and Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, 11:1317–29.

Wolpert, D. M., Miall, R. C., and Kawato, M. (1998). Internal models in the cerebellum. *Trends in Cognitive Sciences*, 2(9):338–47.

Wurtz, R. H. (2008). Neuronal mechanisms of visual stability. *Vision Research*, 48:2070–89.

Xing, J. and Andersen, R. A. (2000a). Memory activity of LIP neurons for sequential eye movements simulated with neural networks. *Journal of Neurophysiology*, 84(2):651–65.

Xing, J. and Andersen, R. A. (2000b). Models of the posterior parietal cortex which perform multimodal integration and represent space in several coordinate frames. *Journal of Cognitive Neuroscience*, 12(4):601–14.

Zhang, P.-Y., Lü, T.-S., and Song, L.-B. (2005). RBF networks-based inverse kinematics of 6R manipulator. *The International Journal of Advanced Manufacturing Technology*, 26(1-2):144–7.

Ziesche, A. and Hamker, F. H. (2011). A computational model for the influence of corollary discharge and proprioception on the peri-saccadic mislocalization of briefly presented stimuli in complete darkness. *The Journal of Neuroscience*, 31:17392–405.

Ziesche, A. and Hamker, F. H. (2014). Brain circuits underlying visual stability across eye movements - converging evidence for a neuro-computational model of area LIP. *Frontiers in Computational Neuroscience*, 8(25).

Zipser, D. and Andersen, R. A. (1988). A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature*, 331(6158):679–84.